Lifshitz tails on the Bethe lattice: a combinatorial approach

Victor Bapst[1] and Guilhem Semerjian[1]

[1]LPTENS, Unité Mixte de Recherche (UMR 8549) du CNRS et de l'ENS,
associée à l'UPMC Univ Paris 06, 24 Rue Lhomond, 75231 Paris Cedex 05, France.

The density of states of disordered hopping models generically exhibits an essential singularity around the edges of its support, known as a Lifshitz tail. We study this phenomenon on the Bethe lattice, i.e. for the large-size limit of random regular graphs, converging locally to the infinite regular tree, for both diagonal and off-diagonal disorder. The exponential growth of the volume and surface of balls on these lattices is an obstacle for the techniques used to characterize the Lifshitz tails in the finite-dimensional case. We circumvent this difficulty by computing bounds on the moments of the density of states, and by deriving their implications on the behavior of the integrated density of states.

I. INTRODUCTION

The seminal work of Anderson [1] has given birth to a vast body of literature on the properties of transport in random environments, both in physics (see [2] for a recent review) and in mathematics (monographs include for instance [3–5]). One central question in this domain is to determine whether a particle can diffuse freely in the environment, according to the dimensionality of the model, the intensity of the disorder, and the energy of the particle. This question can be rephrased in terms of the spreading of eigenvectors of the Hamiltonian corresponding to this energy, or more mathematically in terms of the nature (i.e. absolutely continuous versus pure-point) of its spectrum. Another direction of investigation of these models concerns the density of states [6] of such random Hamiltonians, roughly speaking the distribution of eigenvalues, irrespectively of the nature (localized or extended) of the corresponding eigenvectors. Even though this quantity does not reflect directly the localization properties of the Hamiltonian [7, 8], its study by Lifshitz [9] revealed very early an interesting behavior around the edge of the spectrum, namely a very fast vanishing following an essential singularity known as a Lifshitz tail. This behavior can be explained intuitively as follows. In dimension d, an eigenvector of the non-disordered Hamiltonian with an eigenvalue close to the band edge, say at a very small distance δ from it, is supported on a volume of order $\delta^{-d/2}$. In presence of disorder of bounded amplitude, this vector can give rise to an eigenvalue at an energy δ lower than the (displaced) band edge $E_{\max}$ only if the random potentials on all sites in this volume are close to their maximal value. The probability of this event is exponentially small in the number of involved sites, hence the form of the Lifhsitz tail in dimension d, $\rho(E_{\max}-\delta) \approx \exp[-\delta^{-d/2}]$. This heuristic reasoning can be turned into a rigorous derivation, see for instance chapter VI.2 in [3] and [10] for an illuminating exposition. Note also that even if the density of states is not directly related to the localization aspects of the problem, proofs of localization in finite dimension based on the multi-scale analysis [11] rely crucially on its estimates.

It has been realized early on [12, 13] that the Anderson model could also be studied on the Bethe lattice, thus enabling a mean-field analysis of the localization transition (that does not exist on the fully-connected, complete graph). Since then this version of the model has been the subject of several works both in physics [14–19] and in mathematics [20–27]. Among other results these papers contain numerical procedures to compute the density of states and the location of the localization transition [12, 13, 16–19], as well as proofs of the existence of an absolutely continuous part of the spectrum at low disorder [23–25], and of localization at large disorder or on the border of the spectrum [22]. The study of sparse random matrices [28–34] is actually closely connected to the Anderson problem on the Bethe lattice, even though the perspective taken is slightly different; in the latter case the disorder appears through the connectivity properties of the underlying random graph with fluctuating degrees.

We shall focus in this paper on the Lifshitz tail phenomenon for the Bethe lattice geometry. Heuristic arguments put forward for instance in [14, 19] and a rigorous analysis for a particular type of disorder [35, 36] suggest an even more violent behavior of the density of states in this regime, namely a vanishing of the form $\rho(E_{\max}-\delta) \approx \exp[-\exp[\delta^{-1/2}]]$, that makes this regime very hard to study numerically. As a first example we show in Fig. 1 a plot of the density of states obtained by a standard numerical procedure recalled in Appendix A, which displays an apparent band edge far from its exactly known value. This different form (doubly-exponential) of the Lifshitz tail with respect to the finite-dimensional case can be associated to the exponential (instead of polynomial) growth of the volume of a ball as a function of its radius on the Bethe lattice (formally corresponding to $d \to \infty$). In addition the surface of a ball is asymptotically equivalent to its volume on such a non-amenable graph. This strongly complicates the transposition of the scheme of proofs of Lifshitz tails from the finite-d case to the Bethe lattice one, and indeed rigorous results on the Lifshitz tail behavior on Bethe lattices or sparse random graphs are restricted to sub-critical percolation models [35–37]. The alternative method developed in this paper to circumvent this difficulty consists in studying the asymptotic

behavior of the moments of the density of states. The bounds on the moments that we obtain, and their consequences for the integrated density of states, are in agreement with the doubly-exponential form of the Lifshitz tail on the Bethe lattice [14, 19, 35, 36].

The rest of the paper is organized as follows. In Sec. II we define more precisely the models under study and we state our main results. In Sec. III we collect several expressions of the moments of the density of states. The following two sections (IV and V) present the proofs of our results for the off-diagonal and diagonal disorder case, respectively. Each of these two sections is divided into two subsections focusing on lower and upper bounds on the moments of the density of states. Finally we draw our conclusions and propose perspectives for future work in Sec. VI. More technical aspects of our work are collected in a series of Appendices.

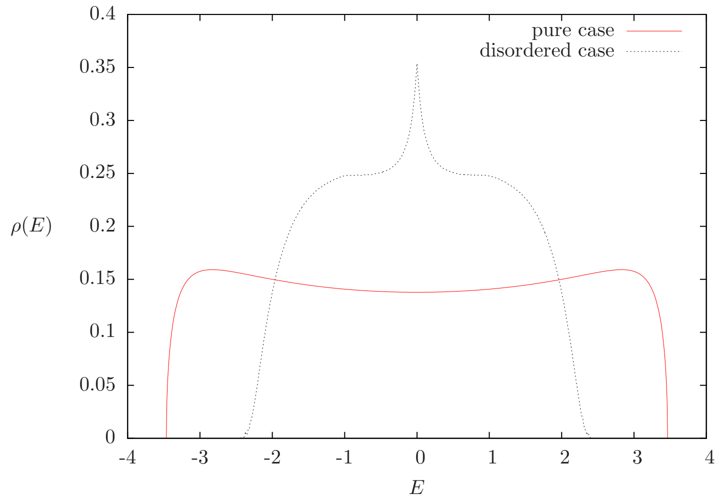


FIG. 1: Density of states on the Bethe lattice of degree $k+1=4$, in the pure ($J_{ij}=1$) and disordered case (with off-diagonal disorder J_{ij} uniformly distributed on $[-1,1]$, see below for a precise definition). The pure case result is the Kesten-McKay [38, 39] density given in Eq. (3). The curve of the disordered case was obtained following the numerical procedure recalled in Appendix A. The support of these two distributions is the same, despite the apparent vanishing of the disordered one due to the strong Lifshitz tail effect.

II. DEFINITIONS AND RESULTS

In this section we shall define the density of states of Anderson models on the Bethe lattice, and state our main results on the asymptotic growth of its moments and their counterparts on the decay of the integrated density of states near the edge of the spectrum. The (integrated) density of states is usually defined, for finite-dimensional models considered on the $\mathbb{Z}^d$ lattice, via a limiting procedure from finite boxes $[-L,L]^d$ with $L\to\infty$. The corresponding construction on an infinite regular tree, namely cutting a depth L neighborhood around an arbitrarily chosen vertex, is not satisfactory because the number of vertices on the surface of a ball of radius L is of the same order as the one of its interior. Finite-size regularizations of the Bethe lattice are instead provided by random regular graphs (see for instance [40] for such a discussion in the context of spin-glasses). We shall hence begin this section by defining both finite (in Sec. II A) and infinite (in Sec. II B) versions of the Bethe lattice models, and highlight the connection between the two. We will in particular define the density of states and its moments, and discuss its regularity properties in Sec.II C. Our main results on the asymptotic growth of the moments of the density of states are given in II E; to make them more intuitively understandable we discuss just before, in Sec. II D, how the behavior of a probability density close to its edge is reflected in the type of growth of its moments.

A. Anderson models on random regular graphs

Let us consider a finite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ on N vertices, and a $N \times N$ symmetric matrix H defined by its matrix elements

$$H_{ij} = \begin{cases} V_i & \text{if } i = j \\ J_{ij} & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} . \tag{1}$$

The real numbers V_i (resp. $J_{ij} = J_{ji}$) represent the influence of the disorder on the vertex i (resp. on the edge between i and j). In the pure case (without disorder), i.e. when $V_i = 0$ and $J_{ij} = 1$ for all vertices and edges, H is nothing but the adjacency matrix of the graph. This well-known case will be considered for comparison purposes in the following.

For a given realization of the Hamiltonian matrix H its empirical spectral measure is defined as a sum of Dirac atoms on its eigenvalues $\lambda_1, \dots, \lambda_N$:

$$\widehat{\rho}_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} . \tag{2}$$

We turn this measure into a random object by taking for $\mathcal{G}$ a random $k+1$-regular graph (i.e. choosing it uniformly at random among all graphs on N vertices, where all vertices have degree $k+1$), for the diagonal elements V_i a set of independent identically distributed (i.i.d.) random variables, and another set of i.i.d. random variables for the J_{ij}'s with $i < j$. The probability measure associated to this construction will be denoted $\mathbb{P}[\cdot]$, and corresponding averages as $\mathbb{E}[\cdot]$. In the following we will concentrate mostly on two particular cases:

- the off-diagonal (or bond) disorder case, in which all the V_i vanishes.
- the diagonal (or site) disorder case, in which all the J_{ij} are equal to 1.

In the former case we shall make the following assumptions on the random edge couplings:

- (OD1) J has a support included in $[-1, 1]$, with $|J|$ not always equal to 1, but which takes values arbitrarily close to 1 with positive probability. In technical terms, we assume the existence of $J_{\min} > 0$ such that $0 < \mathbb{P}[|J| \geq J_0] < 1$ for all J_0 with $J_{\min} \leq J_0 < 1$.
- (OD2) $\forall \alpha > 0$, $a \in (0, 1/2)$, $|\ln \mathbb{P}(J \geq 1 - \epsilon)| e^{-\alpha/\epsilon^a} = O(\epsilon)$ when $\epsilon \to 0$.

The corresponding assumptions for the random energies V_i in the latter case will be:

- (D1) V has a support included in $[0, W]$, distinct from $\{W\}$, giving positive weight to a neighborhood of W.
- (D2) $\forall \alpha > 0$, $a \in (0, 1/2)$, $|\ln \mathbb{P}(V \geq W - \epsilon)| e^{-\alpha/\epsilon^a} = O(\epsilon)$ when $\epsilon \to 0$.

The simplest examples of random variables satisfying these assumptions are the uniformly random ones (on $[-1, 1]$ for J and on $[0, W]$ for V), or the Bernoulli random variable which corresponds, in the off-diagonal case, to the bond percolation model on the Bethe lattice. The assumptions (OD2) and (D2) imply that the occurrence of a J_{ij} (or V_i) close to its maximal value is not by itself a very rare event, otherwise the Lifshitz tail phenomenon would be controlled by single site large deviations and would not be a collective effect anymore, as will become clear in Sec. IV A and Sec. V A. Note that the assumption (D1) does not hold for the Cauchy and Gaussian distributions of diagonal disorder studied for instance in [16, 21, 26, 27]. The choice of a positive support for the diagonal disorder will greatly simplify the proofs, without loss of generality: adding a constant value to the V_i's only shifts the support of $\widehat{\rho}_N$.

The local convergence of random regular graphs to infinite trees ensures [41] that $\widehat{\rho}_N$ converges (weakly, in distribution) in the $N \to \infty$ limit to a probability measure ρ, that we shall call, following the physics convention with a slightly abusive terminology discussed in the next subsections, the density of states. For instance in the pure case where H is the adjacency matrix of a random $(k+1)$-regular graph, ρ is the Kesten-McKay measure [38, 39], supported on $[-2\sqrt{k}, 2\sqrt{k}]$ with the density

$$\rho_{\rm KM}(E) = \frac{k+1}{2\pi} \frac{\sqrt{4k - E^2}}{(k+1)^2 - E^2} . \tag{3}$$

In the next subsection we shall see how to define the measure ρ directly on an infinite object without this $N \to \infty$ limit. Before that we define the average moments of the empirical spectral measure,

$$u_{n,N} \equiv \mathbb{E}\left[\int \lambda^n \mathrm{d}\widehat{\rho}_N(\lambda)\right] = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N} \lambda_i^n\right] = \frac{1}{N}\mathbb{E}\left[\mathrm{tr} H^n\right] = \mathbb{E}\left[(H^n)_{00}\right] \ , \tag{4}$$

where in the last step we have used the invariance of the random graph ensemble with respect to the permutations of the vertices and denoted 0 the index of an arbitrary vertex in V. We shall use the symbol $u_{n,N}$ for a generic disorder distribution, and denote instead these average moments $c_{n,N}$ (resp. $d_{n,N}$) for the off-diagonal (resp. diagonal) disorder case. We define u_n (resp. c_n, d_n) as the $N \to \infty$ limit of $u_{n,N}$ (resp. $c_{n,N}$, $d_{n,N}$). The above stated convergence of $\widehat{\rho}_N$ imply that this limit exists and that the u_n are the moments of the limiting measure, the density of states ρ (a "self-averaging" property). The existence of the limit for the average moments can in fact be obtained in a more direct way: expanding $(H^n)_{00}$ in the last expression of Eq. (4) shows that $u_{n,N}$ depends only on the disorder in the subgraph neighboring the reference vertex 0 within a distance at most $n/2$, which converges in the $N \to \infty$ limit to a regular tree.

B. Anderson models on infinite trees

It is also possible to define Anderson models directly on infinite graphs, that we keep denoting $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The infinite counterpart of the matrix H becomes an operator acting on the elements φ of the Hilbert space $\mathcal{H} = \ell^2(\mathcal{V})$ as

$$(H\varphi)_i = \sum_{j \in \partial i} J_{ij}\, \varphi_j + V_i\, \varphi_i \ , \tag{5}$$

where ∂i stands for the set of neighbors of i in the graph $\mathcal{G}$. When the variables $J_{ij} = J_{ji}$ on the edges of $\mathcal{E}$, the V_i's on the vertices, and the degrees $|\partial i|$ of the vertices, are all uniformly bounded, then H is a self-adjoint operator. For the convenience of the physicist reader who wants to immerse him/herself in the mathematical literature we shall sketch in an informal way some of the concepts used in the mathematical literature on localization, even though this is not the main subject of the paper. Exhaustive reviews of this subject can be found in the monographs [3–5] and in the lecture notes [42].

A major difficulty in dealing with infinite dimensional Hilbert space is that some of the solutions of the eigenvector equation $H\varphi = E\varphi$ are not normalizable, i.e. not in $\ell^2(\mathcal{V})$. The spectrum $\sigma(H)$ has thus to be defined as the complement of the resolvent set. The latter is the set of complex numbers z such that the resolvent operator (or Green function) $(H - z\mathbb{I})^{-1}$ exists and is bounded ($\mathbb{I}$ denotes the identity operator). When H is self-adjoint $\sigma(H) \subset \mathbb{R}$. The spectral theorem [43] is the extension to infinite dimensional self-adjoint operators of the diagonalization of Hermitian matrices. It asserts that H can be written as

$$H = \int \lambda \, \mathrm{d}\mu(\lambda) \ , \tag{6}$$

where μ is a projection valued measure providing a resolution of identity: for all (Borel) subsets I of $\mathbb{R}$, $\mu(I)$ is an orthogonal projection. In the finite dimensional case $\mu(I)$ would project on the subspace spanned by eigenvectors associated to eigenvalues $E \in I$. From this projection valued measure one can define the spectral measures, which are real measures associated to elements $\varphi \in \mathcal{H}$ according to $\mu_\varphi(I) = \langle \varphi | \mu(I) | \varphi \rangle$. Any real measure η can be decomposed [44] in three contributions of different types, $\eta = \eta_{\rm pp} + \eta_{\rm ac} + \eta_{\rm sc}$, where $\eta_{\rm pp}$ is the pure point part of η (its set of Dirac peaks), $\eta_{\rm ac}$ its absolutely continuous part (which gives no weight to sets of zero Lebesgue measure, and has a density thanks to the Radon-Nikodym theorem), and the remainder $\eta_{\rm sc}$ is the singular continuous contribution. Combining this decomposition of real measures with the definition of the spectral measures μ_φ leads to a partition of the Hilbert space as $\mathcal{H} = \mathcal{H}_{\rm ac} \oplus \mathcal{H}_{\rm pp} \oplus \mathcal{H}_{\rm sc}$, where $\mathcal{H}_i$ is defined, for $i = $ pp, ac, sc, as $\mathcal{H}_i = \{\varphi \in \mathcal{H} \,|\, \mu_\varphi \text{ is purely of type } i\}$. Each of these subspaces is stable under the action of H, one can thus define the three spectra $\sigma_i(H)$ as $\sigma(H \restriction \mathcal{H}_i)$, where $H \restriction \mathcal{H}_i$ denotes the restriction of H to $\mathcal{H}_i$. In general $\sigma_{\rm pp}(H)$, $\sigma_{\rm ac}(H)$ and $\sigma_{\rm sc}(H)$ are not disjoint. The pure-point spectrum contains the "true" eigenvalues, corresponding to localized (normalizable) states, while the absolutely continuous spectrum is associated to generalized eigenvalues and extended states. This classification is reflected in the dynamical evolution of the system: a wavepacket constructed from states in the absolutely continuous spectrum spread out at infinity, while pure-point ones remain localized in a finite volume.

This discussion concerned a given operator H; when the coupling constants J_{ij} and on-site energies V_i are turned in random variables, all the quantities defined above become themselves random. However when the J_{ij} and V_i are independent (or more generally form an ergodic process), it follows from ergodicity that the spectrum $\sigma(H)$ is equal, with

probability 1, to a deterministic set independent of the random variables J_{ij} and V_i. Moreover this set is equal to the spectrum of a (pure) model where all random variables are fixed deterministically to their extremal values. A stronger ergodicity statement is actually true: the spectra $\sigma_{\rm pp}(H)$, $\sigma_{\rm ac}(H)$ and $\sigma_{\rm sc}(H)$ are also independent, with probability 1, of the actual realization of the disorder. One of the main goal of the localization studies is the characterization of these sets, depending on the underlying graph and the distribution of the disorder. In unidimensional models the spectrum is completely pure-point for infinitesimal amounts of diagonal disorder [45], while in large dimensions there exists a mobility edge separating the extended and localized [11, 22] parts of the spectrum.

The "density of states" ρ is defined as the average of the spectral measure associated to an element $\varphi \in \mathcal{H}$ supported on a single arbitrary site $0 \in \mathcal{V}$, $\varphi_i = \delta_{i,0}$, that we shall denote

$$\rho = \mathbb{E}[\langle 0|\mu|0\rangle] \ . \tag{7}$$

Its denomination is a bit misleading from a mathematical point of view: ρ is a probability measure, a priori not absolutely continuous and in consequence not associated to a density (we shall come back to this point shortly afterwards). The mathematical literature calls instead integrated density of states (IDS) the cumulative distribution function $N(E) = \rho((-\infty, E])$ which always exists. In finite dimensions (i.e. for $\mathcal{V} = \mathbb{Z}^d$) the IDS can be constructed as follows: call H^L the restriction of the operator H to a box of volume L^d with some boundary conditions. H^L acts on a finite-dimensional space and can thus be diagonalized. Let $N^L(E)$ be the fraction of its eigenvalues smaller than or equal to E; it can then be proven that N^L converges as L is sent to infinity to the (deterministic) distribution function N defined above directly from the infinite-size operator. Another general result on the IDS is that its support coincides with the (almost sure) spectrum $\sigma(H)$.

Let us now come back to the main subject of the paper and specialize some of the definitions above to the (infinite) Bethe lattice. In that case $\mathcal{G}$ is the regular tree in which every vertex has $k+1$ neighbors, sketched in the left panel of Fig. 2 and denoted $\mathbb{T}_k$ in the following. We shall also use the rooted k-ary tree $\widetilde{\mathbb{T}}_k$, in which there is a distinguished vertex (the root) that has no parent node, and every vertex has exactly k children, see right panel of Fig. 2. We will sometimes also consider $\mathbb{T}_k$ as a rooted tree, choosing an arbitrary vertex 0 as the root, and calling children of a vertex $v \neq 0$ its k neighbors away from the root. We take the $J_{ij} = J_{ji}$ as a collection of i.i.d. random variables on the edges,

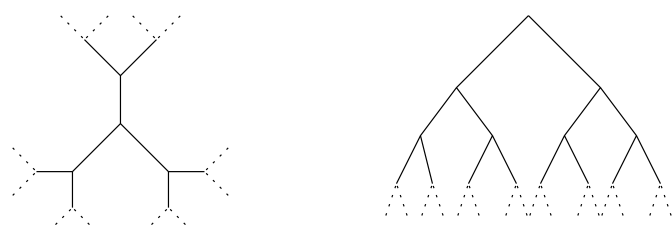


FIG. 2: Left: The regular rooted 3-tree $\mathbb{T}_2$ - Right: The rooted 2-ary tree $\widetilde{\mathbb{T}}_2$.

and similarly with the V_i's another set of i.i.d. random variables. In the pure case ($J_{ij} = 1$, $V_i = 0$) the spectrum of H is the interval $[-2\sqrt{k}, 2\sqrt{k}]$. In the diagonal disorder case ($J_{ij} = 1$) with the V_i i.i.d. random variables supported on $[0, W]$ the ergodicity result mentioned above (see [21] for a discussion of ergodicity on the Bethe lattice) implies that the (almost sure) spectrum of H and the support of ρ is $[-2\sqrt{k}, 2\sqrt{k} + W]$. Similarly in presence of off-diagonal disorder (J_{ij} satisfying (OD1), $V_i = 0$) the support of ρ is the same as in the pure case, $[-2\sqrt{k}, 2\sqrt{k}]$. Trees are in some respects much simpler than finite-dimensional lattices, as they admit a natural recursive decomposition. One can for instance picture $\mathbb{T}_k$ as made of its root, $k+1$ edges from the root to its neighbors, and $k+1$ copies of the rooted k-ary tree $\widetilde{\mathbb{T}}_k$. This decomposition allows to write explicit recursion relations on the diagonal elements of the resolvent (Green function). Let us call $G(z) = \langle 0|(H - z\mathbb{I})^{-1}|0\rangle$ the Green function at the root of $\mathbb{T}_k$, where z has a positive imaginary part. This is a random variable because of the disorder in the J_{ij} and V_i. Using the recursive nature of $\mathbb{T}_k$ and resolvent identities (see App. C for a more explicit derivation with slightly different notations) one can show [23, 41] that $G(z)$ obeys the following recursive distributional equation (RDE):

$$G(z) \overset{\rm d}{=} \frac{1}{V - z - \sum_{i=1}^{k+1} J_i^2 \widetilde{G}_i(z)} \ , \qquad \widetilde{G}(z) \overset{\rm d}{=} \frac{1}{V - z - \sum_{i=1}^{k} J_i^2 \widetilde{G}_i(z)} \ . \tag{8}$$

Here $\overset{\rm d}{=}$ denotes the equality in distribution of random variables, $\widetilde{G}$ is the equivalent of G on $\widetilde{\mathbb{T}}_k$ instead of $\mathbb{T}_k$, and the $\widetilde{G}_i$ (resp. J_i, V) are i.i.d. copies of $\widetilde{G}$ (resp. J and V the distribution of the off-diagonal and diagonal disorder). By definition $\mathbb{E}[G(z)]$ is the Stieltjes transform of ρ, hence the density of the absolutely continuous part of ρ (see the

next subsection for a discussion of the regularity properties of ρ) can be computed from the solution of the RDE as

$$\rho_{\rm ac}(E) = \lim_{\eta \searrow 0} \frac{1}{\pi} {\rm Im} \mathbb{E}[G(E+i\eta)] \ . \tag{9}$$

The equations (8) and (9) can be solved numerically, and this is how we obtained the curves in Figs. 1,3 and 9. We give more details on the numerical procedure in App. A.

As mentioned in the previous section the density of states ρ constructed here directly on the infinite regular tree coincides with the large size limit of the empirical spectral measures for random regular graphs. At variance with finite-dimensional models, the density of states of the Bethe lattice cannot be obtained as the limit of growing subgraphs of the infinite tree: the surface of the latter grows as fast as their volume, which makes this construction ill-behaved.

We defined previously u_n as the n'th moment of ρ. The expression of ρ that follows from Eqs. (6,7) indicates that these moments can be computed on the infinite tree as $u_n = \mathbb{E}[\langle 0 | H^n | 0 \rangle]$. We shall similarly define $\widetilde{u}_n$ as $\mathbb{E}[\langle 0 | H^n | 0 \rangle]$, where H is now restricted on $\widetilde{\mathbb{T}}_k$ rooted at vertex 0, and denote these moments $\widetilde{c}_n$ (resp. $\widetilde{d}_n$) in the off-diagonal (resp. diagonal) case.

Let us note that in principle one could prove analytical results on the Lifshitz tail behavior of the density of states by characterizing directly the solution of the RDE (8). However it is a difficult task to handle quantitatively this kind of equation, this is why we followed the indirect approach via the moments of the density of states in this paper.

C. Absolute continuity of the density of states

As already underlined ρ is in general a probability measure and as such is not guaranteed to be absolutely continuous (i.e. to have a density). However additional assumptions on the random variables J_{ij} and V_i besides (OD1-OD2) or (D1-D2) are known to imply such a regularity of ρ (this is why we denoted $\rho(E)$ instead of $\rho_{\rm ac}(E)$ the density of states in Figs. 1,3 and 9). Let us first discuss the diagonal disorder case. If V has a bounded density, then ρ is absolutely continuous with a bounded density. This is a well-known result for finite-dimensional models, usually called a "Wegner estimate" (see [7] for the original work and [6, 42, 46] for mathematical presentations). The core of the argument in finite dimension goes as follows: the number of eigenvalues of H^L (the regularization of H in a box of size L) below a fixed threshold can only change by one when the potential V_i at one site is varied from its minimal to its maximal value. This is then shown to imply the absolute continuity of the average of N^L, and finally of ρ by taking the $L \to \infty$ limit. The argument can thus be readily extended to the Bethe lattice case thanks to its regularization of finite size N provided by the random regular graphs on N vertices. We stress that the absolute continuity of the density of states should not be mistaken with the nature (localized or extended) of the spectrum for a given realization of the disorder.

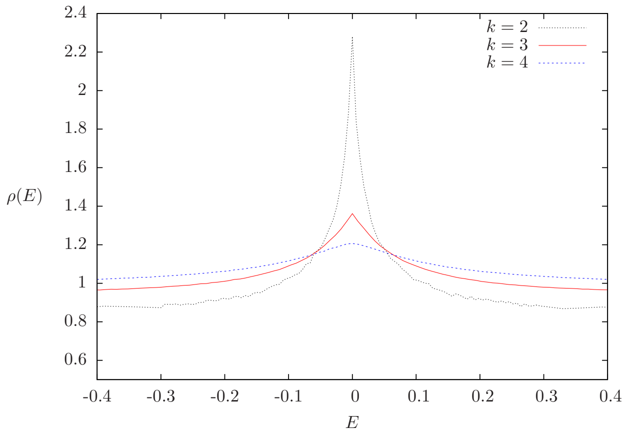
Wegner's argument does not apply directly to the off-diagonal case, but under the additional assumption that J has a Lipshitz continuous density the integrated density of states is locally Lipschitz continuous [47] (hence the density of states is almost everywhere defined and bounded), except possibly in the middle of the band (i.e. for $E=0$). At this energy and in presence of off-diagonal disorder only, the density of states is known to diverge in one dimension [48]. The numerical estimates we obtained for the density of states of the Bethe lattice exhibit a weak non-analyticity around $E=0$ (see Fig. 3). A non-rigorous analysis presented in Appendix A suggests the form $\rho(E) \simeq \rho(0) - \alpha |E|^{\frac{k-1}{2}}$ when $E \to 0$, which is in very good agreement with the numerical results.

D. From a probability density to its moments

In the next subsection we shall present some bounds on the large n behavior of the moments u_n of the density of states ρ, that reflects the behavior of ρ around the edges of its support. To facilitate the intuitive understanding of the bounds we consider here the easy direction of the connection between ρ and u_n, that is we recall how the behavior of a probability distribution at the border of its support influences the growth of its moments.

Let us consider a probability density $\eta(E)$ supported on $[E_-, E_+]$ and its moments $u_n = \int_{E_-}^{E_+} E^n \eta(E) {\rm d}E$. In the large n limit the dominant contribution to the integral arises from regions closer and closer to the edges E_- and E_+, and thus the growth of the moments is controlled by the behavior of $\eta(E)$ close to the edge E_- or E_+ which is largest in absolute value. To simplify the discussion let us assume that η is symmetric and denote $E_0 = E_+ = -E_-$. Then one easily finds the following correspondences (n is implicitly even below):

- If $\eta(E) \sim (E_0 - E)^\alpha$ ($\alpha > 0$), then $\frac{1}{n} \ln u_n = \ln(E_0) - (\alpha+1) \frac{\ln n}{n} + o\left(\frac{\ln n}{n}\right)$.



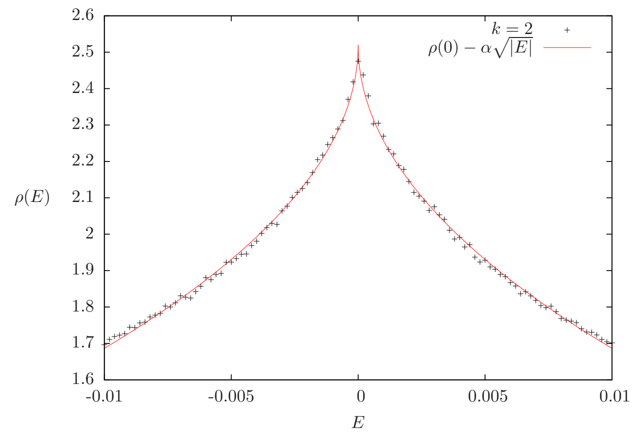


FIG. 3: Left: density of states around $E = 0$ in the off-diagonal disorder case for $k = 2,\ 3,\ 4$. To allow for comparison between these different values of k the couplings have been rescaled, and taken uniformly distributed on $[-\sqrt{2/k}, \sqrt{2/k}]$. Right: detail of the cusp around $E = 0$ for $k = 2$ and a fit to the analytic form derived heuristically in Appendix A.

- If $\eta(E) \sim e^{-\beta(E_0-E)^{-\alpha}}$ $(\alpha, \beta > 0)$, then $\frac{1}{n}\ln u_n = \ln(E_0) - c\,n^{-\frac{1}{\alpha+1}} + o(n^{-\frac{1}{\alpha+1}})$, with c a constant depending on α and β.
- If $\eta(E) \sim e^{-\gamma e^{\beta(E_0-E)^{-\alpha}}}$ $(\alpha, \beta, \gamma > 0)$, then $\frac{1}{n}\ln u_n = \ln(E_0) - \frac{\beta^{1/\alpha}}{E_0}\frac{1}{(\ln n)^{1/\alpha}} + o\left(\frac{1}{(\ln n)^{1/\alpha}}\right)$.

In all cases the value E_0 of the edge of the support controls the dominant (exponential) growth of u_n, that is E_0^n, while the behavior of η in the neighborhood of E_0 yields the subdominant corrections to $(\ln u_n)/n$. From one case to the next in these three examples $\eta(E)$ vanishes faster and faster as $E \to E_0$, and in consequence the corrections to $(\ln u_n)/n$ decay more and more slowly as $n \to \infty$. Note that in the last two cases the vanishing of ρ is fast enough for the integrated density of states $N(E)$ to behave in an equivalent way at the leading order.

The Kesten-McKay distribution of Eq. (3) falls in the scope of the first case with $E_0 = 2\sqrt{k}$ and $\alpha = 1/2$, so that $\frac{1}{n}\ln u_n - \ln(2\sqrt{k}) \sim -\frac{3}{2}\frac{\ln n}{n}$ (which agrees with the exact expression for the moments obtained by McKay [39], recalled in Sec. III D).

As explained in the introduction one expects a doubly-exponential form of the Lifshitz tail for the density of states on Bethe lattices, corresponding to the third case above. This has been established in a rigorous way in [35, 36] for sub-critical bond percolation on the Bethe lattice, under the form

$$\lim_{\delta\to 0}\frac{\ln\ln|\ln(1 - N(E_0 - \delta))|}{\ln\delta} = -\frac{1}{2}\ , \tag{10}$$

where $N(E)$ is the integrated density of states. Its derivative $\rho(E)$ should behave in the same way (this has been proven in finite dimensions in [49]), and in consequence one should expect the correction terms in $(\ln u_n)/n$ to be of order $1/(\ln n)^2$. More precisely, the heuristic reasoning presented in [14, 19] suggests (for off-diagonal disorder) $\rho(E) \sim (2\sqrt{k} - E)^{\left(\frac{k+1}{k}\right)k^{\pi k^{1/4}(2\sqrt{k}-E)^{-1/2}}}$, hence one can apply the formula above with $E_0 = 2\sqrt{k}$, $\alpha = 1/2$ and $\beta = \pi k^{1/4}\ln k$ and make also a prediction for the coefficient of the correction term, $\frac{1}{n}\ln c_n - \ln(2\sqrt{k}) \sim -\frac{(\pi\ln k)^2}{2}\frac{1}{(\ln n)^2}$. The prediction of [19] for diagonal disorder has the same form with $E_0 = 2\sqrt{k} + W$, $\alpha = 1/2$ and $\beta = \pi k^{1/4}\ln k$, which translates into $\frac{1}{n}\ln d_n - \ln(2\sqrt{k} + W) \sim -\frac{(\pi\ln k)^2\sqrt{k}}{2\sqrt{k}+W}\frac{1}{(\ln n)^2}$.

E. Bounds on the moments and the integrated density of states

Let us announce here our main results, the proofs being postponed to sections IV and V. We recall that the random variables defining the model are assumed to follow the assumptions (OD1-OD2) in the off-diagonal disorder case (in particular J_{ij} has support $[-1, 1]$) and (D1-D2) in the diagonal disorder case (V_i is supported on $[0, W]$).

a. Off-diagonal disorder *For all ϵ, $x > 0$, there exists n_0 such that for all even $n > n_0$:*

$$-(1+\epsilon)\frac{(\pi\ln k)^2}{2}\frac{1}{(\ln n)^2} \le \frac{1}{n}\ln c_n - \ln(2\sqrt{k}) \le -\frac{1}{(\ln n)^x}\frac{(\pi\ln k)^2}{2}\frac{1}{(\ln n)^2}\ . \tag{11}$$

b. Diagonal disorder For all ϵ, $x > 0$, there exists n_0 such that for all $n > n_0$:

$$-(1+\epsilon)\frac{(\pi \ln k)^2}{2}\frac{2\sqrt{k}}{(2\sqrt{k}+W)}\frac{1}{(\ln n)^2} \leq \frac{1}{n}\ln d_n - \ln(2\sqrt{k}+W) \leq -\frac{1}{(\ln n)^x}\frac{(\pi \ln k)^2}{2}\frac{2\sqrt{k}}{(2\sqrt{k}+W)}\frac{1}{(\ln n)^2} \ . \tag{12}$$

Let us first precise the level of rigor of these results: the proofs of the lowerbounds given in Sec. IV A and Sec. V A are mathematically rigorous. The upperbounds rely on an explicit computation that we did not turn in a rigorous derivation, but that we checked numerically with a very high accuracy, as discussed more precisely in Sec. IV B and Sec. V B.

Note also that in the statement of the upperbounds the multiplicative constant is actually irrelevant because of the condition $x > 0$; we chose however to write them in this suggestive form because we conjecture that the leading term of the asymptotic expansion corresponds to $\epsilon = x = 0$, i.e. that

$$\frac{1}{n}\ln c_n = \ln(2\sqrt{k}) - \frac{(\pi \ln k)^2}{2}\frac{1}{(\ln n)^2} + o\left(\frac{1}{(\ln n)^2}\right) \ , \tag{13}$$

and a similar conjecture in the diagonal case, in full agreement with the heuristic prediction on ρ discussed above.

We now give the implication of these bounds on the moments for the behavior of the density of states itself. The support of the latter being bounded, ρ is unambiguously determined by the knowledge of all its moments [44]. The difficulty here is that we only have an asymptotic control on the moments, and that the very slow decay of their corrections hampers the use of transfer (tauberian) theorems [44]. It is however possible to turn the above statements on the moments of ρ into bounds on the integrated density of states $N(E) = \rho([-\infty, E])$:

c. Off-diagonal disorder For all ϵ, $x > 0$, and δ small enough:

$$e^{-k^{\pi k^{1/4}\sqrt{\frac{(1+\epsilon)}{\delta}}}} \leq N(-2\sqrt{k}+\delta) = 1 - N(2\sqrt{k}-\delta) \leq e^{-k^{\pi k^{1/4}\left(\frac{1}{\delta}\right)^{\frac{1}{2+x}}}} \tag{14}$$

d. Diagonal disorder For all ϵ, $x > 0$, and δ small enough:

$$e^{-k^{\pi k^{1/4}\sqrt{\frac{(1+\epsilon)}{\delta}}}} \leq 1 - N(2\sqrt{k}+W-\delta) \leq e^{-k^{\pi k^{1/4}\left(\frac{1}{\delta}\right)^{\frac{1}{2+x}}}} \tag{15}$$

The proofs of these two statements are deferred to Appendix B. As mentionned in II A, the positivity of the support of V in hypothesis (D1) simplifies the derivation of the bounds on the moments, but is not needed for (15) to hold, this last statement only requiring that V is bounded and gives positive weight to the neighborhood of its upperbound W. Also note that as before the constants on the right-hand sides above are irrelevant due to the freedom in the choice of x, and are only here to suggest what we believe should be the asymptotic of $\ln|\ln(1-N(E_0-\delta))|$, with $E_0 = 2\sqrt{k}$ in the off-diagonal disorder case and $2\sqrt{k}+W$ in the diagonal disorder case. In fact, a more modest and concise statement is:

$$\lim_{\delta\to 0}\frac{\ln\ln|\ln(1-N(E_0-\delta))|}{\ln \delta} = -\frac{1}{2} \ , \tag{16}$$

In particular this extends the result of [35, 36] on the bond percolation model to the percolating phase.

Finally let us emphasize that in the course of the proof of the lower bounds we shall obtain in a rigorous way that

$$\lim_{n\to\infty}\frac{1}{n}\ln c_n = \ln(2\sqrt{k}) \ , \qquad \lim_{n\to\infty}\frac{1}{n}\ln d_n = \ln(2\sqrt{k}+W) \ , \tag{17}$$

with n even in the first case. This is easily seen to imply (see [39] or App. B for more details) that the support of the density of states of the off-diagonal (resp. diagonal) disorder model extends up to $2\sqrt{k}$ (resp. $2\sqrt{k}+W$), as expected from the ergodicity arguments discussed above.

III. VARIOUS EXPRESSIONS OF THE MOMENTS

A. Walks on the tree

From the definition of the model on the (either finite or infinite) Bethe lattice it should be clear that the average moments u_n (this notation encompassing both the diagonal and off-diagonal disorder cases) can be written as:

$$u_n = \sum_{i_1,i_2,\ldots,i_{n-1}} \mathbb{E}[H_{0i_1}H_{i_1i_2}\ldots H_{i_{n-1}0}] \ . \tag{18}$$

This is a sum over (lazy) walks of lengths n on $\mathbb{T}_k$, starting and ending at its root 0, i.e. over sequences $0 = i_0, i_1, \dots, i_{n-1}, i_n = 0$ of vertices of $\mathbb{T}_k$, such that for all $j \in [0, n-1]$ either $i_j = i_{j+1}$ or (i_j, i_{j+1}) is an edge of $\mathbb{T}_k$. In the former case, that we shall call a self-bond step around vertex i_j, the matrix element is $H_{i_j,i_{j+1}} = V_{i_j}$. In the latter case $H_{i_j,i_{j+1}} = J_{i_j,i_{j+1}}$. As trees have no cycle each edge is visited an even (possibly null) number of times during a closed walk. We denote $\mathcal{W}_n$ the set of walks of length n around the root 0, and for a walk $\omega = (i_1, \dots, i_{n-1}) \in \mathcal{W}_n$ we define its weight π as $\pi(\omega) \equiv \mathbb{E}[H_{0i_1} H_{i_1 i_2} \dots H_{i_{n-1}0}]$, so that $u_n = \sum_{\omega \in \mathcal{W}_n} \pi(\omega)$. A more explicit expression of π is obtained by defining $n_e(\omega)$ as half the number of times the edge $e \in \mathcal{E}$ is crossed during the walk ω, and $s_v(\omega)$ the number of self-bond steps around vertex $v \in \mathcal{V}$. Indeed, as the disorder is given by i.i.d. random variables,

$$\pi(\omega) = \prod_{e \in \mathcal{E}} \mathbb{E}[J^{2n_e(\omega)}] \prod_{v \in \mathcal{V}} \mathbb{E}[V^{s_v(\omega)}] \ . \tag{19}$$

In the off-diagonal disorder case only walks without self-bond steps have a non-zero weight; to avoid confusion we shall thus denote in this case $\mathcal{M}_n$ the set of walks of length n without self-bond steps.

It will be useful in the following (in particular in the proofs of the upperbounds in Sec. IV B and V B) to refine the description of a walk, and to partition the set of walks $\mathcal{W}_n$ in various subsets. We define the support $\sigma(\omega)$ of a walk ω as the set of edges of the tree visited at least once by ω, i.e. $\sigma(\omega) = \{e \in \mathcal{E} | n_e(\omega) \geq 1\}$, and the size of a support as the number of edges it contains. A support of size r is a subtree of $\mathbb{T}_k$ containing the root and r edges, i.e. a tree of r edges where the root has at most $k+1$ children, and all other vertices have at most k children. We shall denote $\mathcal{S}^r$ the set of the supports of size r.

The skeleton $\widehat{\sigma}(\omega)$ is defined as the support of ω, supplemented by the numbers $\{n_e(\omega)\}_{e \in \sigma(\omega)}$. More generally a skeleton $\widehat{\sigma}$ is made of a support σ and a set of positive integers $\{n_e\}_{e \in \sigma}$; we call $2 \sum_{e \in \sigma} n_e$ the length of the skeleton $\widehat{\sigma}$, and denote $\widehat{\mathcal{S}}^r_n(\sigma)$ the set of skeletons of length n based on a support σ of size r.

We will also call self-support of a walk ω the set of vertices around which at least one self-bond step is taken, i.e. the set $\{v \in \mathcal{V} | s_v(\omega) \geq 1\}$.

The construction of a walk $\omega \in \mathcal{M}_n$, i.e. of length n without self-bond, amounts thus to the successive choices of:

- a support σ of size $r \leq \frac{n}{2}$.
- a positive integer n_e for each edge e of the support, such that $2 \sum_{e \in \sigma} n_e = n$. This completes the choice of the skeleton of the walk.
- an ordering of the visited edges, that is, a mapping φ from $\{1, \dots, n\}$ to σ compatible with the tree structure and the n_e. In technical terms, one must have $\forall e \in \sigma, |\varphi^{-1}(\{e\})| = 2n_e$ and $(\varphi(1), \dots, \varphi(n))$ must correspond to a walk on the vertices covered by σ.

Note that there are in general several walks compatible with a given skeleton, as explained on a simple example in Fig. 4; we shall come back to this issue in Sec.IV B-IV C.

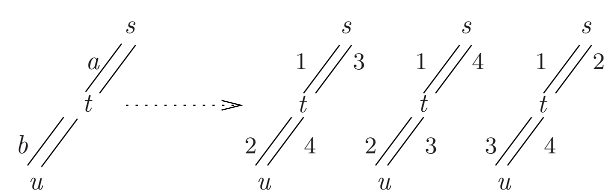


FIG. 4: A skeleton $\widehat{\sigma}$ of size 2 and length 8, made of the support $\sigma = \{a, b\}$ and the number of crossings of these two edges, $n_a = 2$ and $n_b = 2$. The three vertices involved are denoted $\{s = 0, t, u\}$, on the right are drawn the three walks compatible with the skeleton (i.e. $\omega = (s, t, u, t, s, t, u, t, s)$, $\omega = (s, t, u, t, u, t, s, t, s)$ and $\omega = (s, t, s, t, u, t, u, t, s)$), the numbers close to the edges being the order in which they are crossed from top to bottom.

.

The decomposition given above can be generalized to the case where self-bond steps are allowed. Indeed, the definition of such a walk corresponds then to choose:

- a number $s \in [0, n]$ of self-bond steps (such that $n - s$ is even).
- a skeleton of length $n - s$.
- an ordering of the edges of the support, satisfying the same constraints as given above in absence of self-bond steps. This yields a walk $\overline{\omega} \in \mathcal{M}_{n-s}$.

- s times $0 \leq t_1 \leq t_2 \leq \dots t_s \leq n-s$, specifying that self-bond steps have to be inserted after $t_1, t_2, \dots, t_s$ steps of $\overline{\omega}$.

B. A recursive computation

1. Recursion relation

Taking advantage of the recursive structure of $\mathbb{T}_k$ one can set up a recursive computation of u_n. In this section we present the equations and their intuitive interpretation, their formal derivation (which is in essence a series expansion of the recursion (8) between the Green functions) being deferred to App. C. Similar recursive equations on the moments have been derived in [30, 32] for sparse random matrices, where the randomness lies in the degrees of the vertices.

Consider a closed walk of length n starting at the root 0 of $\mathbb{T}_k$, call s the number of self-bond steps made at 0 and p_i half the number of times the edge between the root and its neighbor $i \in [1, k+1]$ is crossed. The walk will make a number m_i of steps in each of the subtrees $\mathbb{T}^{(i)}$, the copy of $\widetilde{\mathbb{T}}_k$ rooted at i, such that $s + \sum_i (m_i + 2p_i) = n$. One must however take into account the fact that these m_i steps are to be divided in p_i closed walks on $\mathbb{T}^{(i)}$, separated by a visit to the root 0. In each of these closed walks the disorder encountered in $\mathbb{T}^{(i)}$ is the same, hence the contribution of the various closed walks on $\mathbb{T}^{(i)}$ are correlated. Summarizing these observations one obtains the following expression of u_n,

$$u_n = \sum_{\substack{s,p_1,m_1,\dots,p_{k+1},m_{k+1}\geq 0 \\ s+m_1+2p_1+\dots+m_{k+1}+2p_{k+1}=n}} u(m_1,p_1)\dots u(m_{k+1},p_{k+1}) \binom{s+p}{s,p_1,\dots,p_{k+1}} \mathbb{E}[J^{2p_1}]\dots\mathbb{E}[J^{2p_{k+1}}]\mathbb{E}[V^s] \ , \tag{20}$$

where in the multinomial coefficient we denoted $p = p_1 + \dots + p_{k+1}$. This coefficient counts the number of orderings of the steps made each time the walk steps out the root, either for a self-bond step or towards one of the $k+1$ neighbors. The quantity $u(m,p)$ is the average contribution of the steps made in one of the subtrees $\mathbb{T}^{(i)}$, when m steps are performed in $\mathbb{T}^{(i)}$, divided in p epochs separated by visits to the ancestor 0. A more explicit interpretation of $u(m,p)$ is the following: introduce an additional vertex (-1) connected only to the root 0 of $\widetilde{\mathbb{T}}_k$ with an edge of weight $J_{-1,0}=1$, and define $u(m,p)$ as the weighted sum of the walks of length $m+2p$ on $\widetilde{\mathbb{T}}_k \cup (-1)$, that start and end at 0 and that are constrained to visit exactly p times the vertex (-1). Following the same reasoning as above one can compute the $u(m,p)$'s by recursion, according to the rules $u(0,q)=1$, $u(n,0)=\delta_{n,0}$,

$$u(n,q) = \sum_{\substack{s,p_1,m_1,\dots,p_k,m_k\geq 0 \\ s+m_1+2p_1+\dots+m_k+2p_k=n}} u(m_1,p_1)\dots u(m_k,p_k) \binom{q-1+s+p}{q-1,s,p_1,\dots,p_k} \mathbb{E}[J^{2p_1}]\dots\mathbb{E}[J^{2p_k}]\mathbb{E}[V^s] \quad \text{for} \quad q \geq 1 \ , \tag{21}$$

where again in the multinomial coefficient p stands for $p_1+\dots+p_k$, and this coefficient counts the number of orderings of the steps taken from the root of $\mathbb{T}^{(i)}$ (there appears $q-1$ because the first passage of the walk on i necessarily arrives from the ancestor). Note that in particular $u(n,1)$ is equal to $\widetilde{u}_n$, the moments of order n for the walks on $\widetilde{\mathbb{T}}_k$.

2. Numerical evaluation

One can check that Eq. (21) does indeed provide a recursive scheme to compute all the $u(n,q)$: the computation of these values (and of the moment u_n from Eq. (20)) at rank n only requires the knowledge of $u(m,p)$ with $m<n$ and $2p \leq n$. The number of terms to sum in order to obtain a new $u(n,q)$ grows as $O(n^{2k})$ ($O(n^{2k-1})$ in the off-diagonal case), and so the computation of u_n requires a time that grows as $O(n^{2k+2})$ ($O(n^{2k+1})$ in the off-diagonal case). Though this is a polynomial time complexity the exponent is high (at least 5), and in practice only rather limited values of n are accessible within a reasonable time on present computers. For illustration we present in Table I numerical results up to $n=62$, in the off-diagonal disorder model with $k=2$, along with the corresponding values for the pure case.

A lowerbound on u_n can also be obtained numerically for larger values of n: as all terms in Eq. (21) are positive, the quantities $u'(n,q)$ obtained according to the recursion (21) in which all sums on p_i are restricted to $p_i \leq p_{\max}$, a threshold fixed independently of n, are smaller than the $u(n,q)$'s. The complexity of the computation of $u'(n,q)$ is reduced to $O(n^{k+1})$ ($O(n^k)$ for the off-diagonal case); for $k=2$ we could reach values of n of the order of 200, using

n	2	4	8	16	24	32	40	50	62
pure	2	8	224	$3.66\ 10^5$	$8.52\ 10^8$	$2.32\ 10^{12}$	$6.88\ 10^{15}$	$1.63\ 10^{20}$	$3.12\ 10^{25}$
disordered	0.67	1.06	5.37	375	$4.67\ 10^4$	$7.93\ 10^6$	$1.66\ 10^9$	$1.64\ 10^{12}$	$8.29\ 10^{15}$

TABLE I: Average value of the moments $\tilde{c}_n = u(n,1)$ on the 2-ary tree $\widetilde{\mathbb{T}}_{k=2}$, in the off-diagonal disordered model with J_{ij} uniformly random on $[-1,1]$. The results of the pure case ($J_{ij}=1$) can be directly computed from Eq. (25).

$p_{\max}=10$. The approximation coming from the finite value of $p_{\max}$ becomes worse and worse as n grows larger. In any case, the values of n reachable numerically remains very far from the regime in which we expect the asymptotic scaling of the moments presented in Sec. II E to hold.

C. Explicit expression

By "unfolding" the recursive equations (20,21), i.e. replacing the $u(n,q)$ in the r.h.s. by their recursive expressions, one obtains a formula for the moment u_n in which the summation over the number of visits of all edges and self-bonds of the tree is made explicit. In order to simplify the notations let us define a labelling on the edges and vertices of the tree by: the root has label 0, the children of a vertex v have indices $v_1,\dots,v_k$ (except the root whose neighbors are denoted $1,\dots,k+1$) and an edge that connects a vertex v to its ancestor v' has index v. With these definitions and the previously introduced notations, we obtain the following expression for u_n:

$$u_n = \sum_{\substack{\{n_v\geq 0\}_{v\in\mathcal{V}\setminus 0}\\ \{s_v\geq 0\}_{v\in\mathcal{V}}}}{}^{'} \binom{s_0+n_1+n_2+\cdots+n_{k+1}}{s_0,n_1,n_2,\dots,n_{k+1}}\mathbb{E}[V^{s_0}]\prod_{v\in\mathcal{V}\setminus 0}\binom{n_v-1+s_v+n_{v_1}+\cdots+n_{v_k}}{n_v-1,s_v,n_{v_1},\dots,n_{v_k}}\mathbb{E}[J^{2n_v}]\mathbb{E}[V^{s_v}]\ , \qquad (22)$$

where the prime on the sum denotes the constraint $s_0+\sum_{v\in\mathcal{V}\setminus 0}(2n_v+s_v)=n$ that has to be verified by the summands. We also used the convention $\binom{\cdots}{-1,s,n_1,\dots,n_k}=\delta_{s,0}\delta_{n_1,0}\dots\delta_{n_k,0}$, which arises from the boundary condition $u(n,0)=\delta_{n,0}$. More explicitly, this convention for the multinomial coefficient enforces the connected character of the set of non-zero elements n_v, which in consequence form a valid skeleton for the walks (incidentally this also implies that the number of non-zero terms in the sum is finite for any finite n). This expression can be thus interpreted as a sum over skeletons and number of self-bond steps of a walk, the product of the multinomial coefficients counting the number of walks compatible with such values of $\{n_e\}_{e\in\mathcal{E}}$, $\{s_v\}_{v\in\mathcal{V}}$, that arises from the freedom of choice in the order the steps around each vertex are taken. Let us remark finally that this expression can easily be generalized to an arbitrary tree, at the price of slightly more cumbersome notations.

D. The moments of the pure case

As mentioned above the density of states is given in the pure case ($J_{ij}=1$, $V_i=0$) by the Kesten-McKay law whose density was recalled in Eq. (3). As a matter of fact this measure was determined by McKay [39] via the computation of its moments. Let us briefly recall this derivation, that we shall extend in Sec. IV A and V A to obtain the lowerbounds on the moments in the disordered case. We denote c_n^0 and $\widetilde{c}_n^0$ the values of the moments on $\mathbb{T}_k$ and $\widetilde{\mathbb{T}}_k$ respectively for this pure case, as well as $u^0(n,q)$ for the solution of Eq. (21).

Consider a closed walk on $\widetilde{\mathbb{T}}_k$ of length n, $\omega=(0=i_0,i_1,\dots,i_{n-1},i_n=0)$, and call depth of a vertex the distance that separates it from the root. Then the sequence $(0=d_0,d_1,\dots,d_{n-1},d_n=0)$ of the depths of the vertices visited by ω forms a Dyck path, i.e. they are non-negative integers, $|d_{i+1}-d_i|=1$, with 0 as final and initial values. This correspondence is illustrated in Fig. 5.

The computation of the number of Dyck paths of length n is a classical exercise in combinatorics [50, 51], which yields the Catalan number $\binom{n}{n/2}\frac{1}{n/2+1}$ for n even, 0 otherwise. Coming back to the computation of $\widetilde{c}_n$, the important point to realize is that there are exactly $k^{n/2}$ walks on $\widetilde{\mathbb{T}}_k$ that give the same Dyck path of length n: at every of the $n/2$ steps taken away from the root there are k possible children to choose from. Hence we obtain in the pure case

$$u^0(n,1)=\widetilde{c}_n^0=\binom{n}{n/2}\frac{1}{n/2+1}k^{n/2} \qquad (23)$$

for n even, zero otherwise.

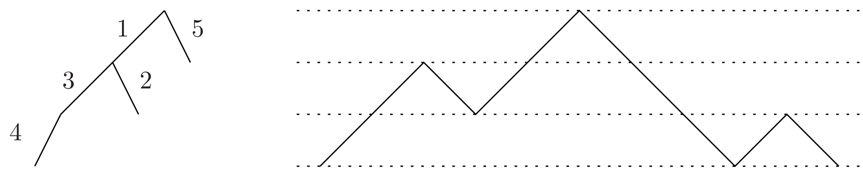


FIG. 5: A walk on the 2-ary tree $\widetilde{\mathbb{T}}_2$ and the corresponding Dyck path. In the left figure the numbers give the order in which the edges are crossed from top to bottom.

A similar computation can also be done on $\mathbb{T}_k$; the depth projection of a closed walk of length n on $\mathbb{T}_k$ still yields a Dyck path of length n. However the number of distinct walks corresponding to the same Dyck paths depends on the number of its returns to the origin. Indeed there are $k+1$ choices for a step out of the root, and only k for steps taken away from the root from another vertex. The number of Dyck paths of length n with i returns to the origin can also be enumerated [50], the result being $\binom{n-i}{n/2}\frac{i}{n-i}$. This yields the moment

$$c_n^0 = \sum_{i=1}^{n/2} \binom{n-i}{n/2} \frac{i}{n-i} (k+1)^i k^{n/2-i} \ , \tag{24}$$

which can be checked to be the n-th moment (for n even) of the Kesten-McKay distribution of Eq. (3). A crude estimation of the asymptotic behavior of c_n^0 shows that $\lim \frac{1}{n} \ln c_n^0 = \ln(2\sqrt{k})$, which reflects the fact that the support of the distribution in Eq. (3) is $2\sqrt{k}$ (recall the discussion in Sec. II D). The same asymptotic behavior is readily found for $\widetilde{c}_n^0$ from Eq. (23).

In the pure case one can actually find the solution of the recursive equation (21) by similar considerations on the depth projections of closed walks,

$$u^0(n,q) = \binom{n+q}{n/2+q} \frac{q}{n+q} k^{n/2} \tag{25}$$

for n even, zero otherwise. We show in Appendix D an explicit verification of this statement.

Let us make a final series of remarks before closing this section:

- the projection from a walk on a tree to the Dyck path representing the distance from the root to the walker is a powerful tool in the pure case because it is easy to count the number of walks corresponding to a given Dyck path. Unfortunately it is much harder to compute the total weight of walks sharing a Dyck path in the (off-diagonal) disorder case: the disorder is kept fixed along the walk, this induces correlations between the choices of the branch followed by the walk which are lost in the projection on its depth.

- one can try to apply a saddle-point estimate to the recursion equation (21): writing $u(n, q = nx) \simeq e^{nf(x)}$ gives a functional equation on $f(x)$ whose solution is found to be, independently of the disorder, the one of the pure case $f(x) = \ln k + (2+x)\ln(2+x) - (1+x)\ln(1+x)$. However, we were not able to get the non-exponential corrections to this formula in the disordered case, as the difference in (21) between the pure and the disordered case is only in polynomials factors, whereas according to II E we expect a slower dominant change in $u(n,q)$. Another difficulty is that the corrections to $u(n,q)$ do not depend much on the particular shape of disorder, whereas (21) seems to. An alternative approach that we did not pursue would be to seek a specific distribution of the disorder such that its moments make Eq. (21) solvable with a closed combinatorial form for $u(n,q)$.

- the asymptotic behavior of the "shifted pure case" where $J_{ij} = 1$, $V_i = W$ a fixed constant, for which the moments are denoted d_n^0, can be obtained either by shifting the density of Eq. (3) to $\rho_{\rm KM}(E-W)$ or by depth projections leading to Motzkin paths (see App. E for a definition) with a weight W assigned to self-bond (horizontal) steps. This yields $\lim \frac{1}{n} \ln d_n^0 = \ln(2\sqrt{k} + W)$, a result which shall be used for comparison in the discussion of the diagonal disorder model.

IV. PROOFS FOR THE OFF-DIAGONAL DISORDER MODEL

This section contains the proofs of the bounds stated in Eq. (11) for the off-diagonal disorder case, i.e. for vertex random variables V_i all vanishing, and random edge couplings J_{ij} drawn in $[-1,1]$ with a probability distribution

satisfying the assumptions (OD1-OD2) of Sec. II A. We denote c_n the moments of the density of states, and assume implicitly below that n is even, as all odd moments obviously vanish in this case, the length of a closed walk on a tree being by definition even.

A. Lower bound

Let us first explain the strategy underlying the proof of the lowerbound, reminiscent of the procedure for finite-dimensional models sketched in the introduction. We shall indeed lowerbound the moments c_n of the density of states by restricting the contributions of walks to those confined to a small volume around the root (at a distance smaller than h), under the condition that the disorder takes a large value in this volume (i.e. all $|J_{ij}|$ shall be greater than J_0). An optimal compromise will then be found between these two restrictions: taking h too small reduces too drastically the possible contributions to c_n, while the probability of having large disorder falls off quickly if h is taken too large. We now turn this reasoning into an explicit derivation.

We recall that $c_n = \sum_{\omega \in \mathcal{M}_n} \pi(\omega)$, where $\mathcal{M}_n$ denotes the set of closed walks (with no self-bond steps) of length n starting at the root of $\mathbb{T}_k$. The weight of a walk $\pi(\omega)$ is given by $\mathbb{E}[\prod_e J_e^{2n_e(\omega)}]$, where e runs over the edges of $\mathbb{T}_k$ and $n_e(\omega)$ is a non-negative integer equal to half the number of times the walk crosses the edge e. As all weights $\pi(\omega)$ are positive, c_n can be lowerbounded by restricting the sum to $\widetilde{\mathcal{M}}_n$, the walks on $\widetilde{\mathbb{T}}_k$, and further to $\widetilde{\mathcal{M}}_{n,h}$, the walks on $\widetilde{\mathbb{T}}_k$ which go at a distance at most h from the root:

$$c_n \geq \widetilde{c}_n = \sum_{\omega \in \widetilde{\mathcal{M}}_n} \pi(\omega) \geq \sum_{\omega \in \widetilde{\mathcal{M}}_{n,h}} \pi(\omega) \ . \tag{26}$$

Let us denote $\mathcal{E}_h$ the set of edges with endpoints at distance at most h from the root of $\widetilde{\mathbb{T}}_k$, M the event: $\{|J_e| \geq J_0 \ \forall e \in \mathcal{E}_h\}$, where J_0 is an arbitrary threshold with $0 < J_0 < 1$, and $\overline{M}$ the complementary event. For an arbitrary walk ω in $\widetilde{\mathcal{M}}_{n,h}$ we have

$$\begin{aligned} \pi(\omega) &= \mathbb{E}\left[\prod_{e \in \mathcal{E}_h} J_e^{2n_e(\omega)} \middle| M \right] \mathbb{P}[M] + \mathbb{E}\left[\prod_{e \in \mathcal{E}_h} J_e^{2n_e(\omega)} \middle| \overline{M} \right] \mathbb{P}[\overline{M}] \\ &\geq \mathbb{E}\left[\prod_{e \in \mathcal{E}_h} J_e^{2n_e(\omega)} \middle| M \right] \mathbb{P}[M] \\ &\geq J_0^n \, \mathbb{P}[|J| \geq J_0]^{|\mathcal{E}_h|} \ , \end{aligned} \tag{27}$$

where in the last step we have used the independence of the J_e on different edges and the fact that $\sum_e 2n_e(\omega) = n$. This lowerbound being independent of ω, it is now enough to control the number $|\widetilde{\mathcal{M}}_{n,h}|$ of walks on $\widetilde{\mathbb{T}}_k$ which remain at a distance smaller or equal to h from the root. Using the projection from a walk to its depth explained in Sec. III D, one realizes that $|\widetilde{\mathcal{M}}_{n,h}| = k^{n/2} m_{n,h}$, where $m_{n,h}$ is the number of Dyck paths of length n and of height at most h. These paths can be enumerated with standard combinatorial techniques [51, 52]; in particular we show in App. E (along with a more precise estimate of its asymptotic behavior) that $m_{n,h}$ obeys the inequality

$$m_{n,h} \geq \left(\frac{2}{h+2} \right)^3 \left(2 \cos\left(\frac{\pi}{h+2} \right) \right)^n \ , \tag{28}$$

valid for all values of h and of (even) n. We can thus write

$$c_n \geq J_0^n \, \mathbb{P}[|J| \geq J_0]^{|\mathcal{E}_h|} \left(\frac{2}{h+2} \right)^3 \left(2\sqrt{k} \cos\left(\frac{\pi}{h+2} \right) \right)^n \ , \tag{29}$$

or equivalently after taking logarithm, subtracting the leading order of the pure case and using the fact that $|\mathcal{E}_h| = \frac{k^{h+1}-k}{k-1} \leq \frac{k^{h+1}}{k-1}$:

$$\frac{1}{n} \ln c_n - \ln(2\sqrt{k}) \geq \ln(J_0) + \frac{k^{h+1}}{n(k-1)} \ln \mathbb{P}[|J| \geq J_0] + \frac{3}{n} \ln\left(\frac{2}{h+2} \right) + \ln \cos\left(\frac{\pi}{h+2} \right) \ . \tag{30}$$

From this equation we shall first show that $\lim_{n\to\infty} \frac{1}{n} \ln c_n = \ln(2\sqrt{k})$. Fixing h and J_0 and letting $n \to \infty$ in the inequality above yields

$$\liminf_{n\to\infty} \frac{1}{n} \ln c_n - \ln(2\sqrt{k}) \geq \ln(J_0) + \ln\cos\left(\frac{\pi}{h+2}\right) \ . \tag{31}$$

Sending now J_0 to 1 and h to infinity we obtain $\liminf \frac{1}{n}\ln c_n \geq \ln(2\sqrt{k})$. On the other hand we obviously have $c_n \leq c_n^0$, the value in the pure case where all $J_{ij} = 1$, and the results of McKay [39] recalled in Sec. III D implies that $\frac{1}{n}\ln c_n^0 \to \ln(2\sqrt{k})$. From these matching lower and upper bounds it follows that $\lim_{n\to\infty} \frac{1}{n}\ln c_n = \ln(2\sqrt{k})$, i.e. that the edge of the density of states is the same as in the pure case, as expected for the ergodicity reasons discussed in Sec. II B. Note that the only assumption on the distribution of J which was necessary here is the existence of some $J_{\rm min} \geq 0$ such that $\mathbb{P}[|J| \geq J_0] > 0$ for all J_0 with $J_{\rm min} \leq J_0 < 1$.

The more precise result stated in Eq. (11) is obtained by taking $h \to \infty$ and $J_0 \to 1$ in an n-dependent way. By inspection of the last term in the r.h.s. of Eq. (30) one realizes that the best bound will be achieved by taking h as large as possible. The limitation on the possible range of h comes instead from the second term, in which k^{h+1} has to be small compared to n. We shall in consequence take $h = \lfloor \alpha \ln n \rfloor$ with $\alpha < 1/\ln k$ and $J_0 = 1 - \frac{1}{(\ln n)^{2+y}}$ where $y > 0$ is an arbitrary positive constant. In the limit $n \to \infty$ the r.h.s. of Eq. (30) becomes

$$-\frac{\pi^2}{2\alpha^2}\frac{1}{(\ln n)^2} + \frac{k}{k-1}\frac{1}{n^{1-\alpha\ln k}} \ln \mathbb{P}\left[|J| \geq 1 - \frac{1}{(\ln n)^{2+y}}\right] + O\left(\frac{1}{(\ln n)^{2+y}}\right) \ . \tag{32}$$

For all values of $\alpha < 1/\ln k$ and $y > 0$ the second term is, thanks to the assumption (OD2) on the random variable J, negligible with respect to the first, which proves the statement of the lowerbound in Eq. (11).

If the assumption (OD2) were violated the dominant effect at the origin of the reduction of the density of states near the edge would be the extremely low probability for a single $|J_{ij}|$ to be close to 1, not the collective effect of all $|J_{ij}|$ in a given volume being large. In that case the form of the lowerbound in Eq. (11) would have to be modified and would depend on the precise form of the density of J around 1. The transition from the collective to the single edge dominated regime of the Lifshitz tail phenomenon, depending on the form of the edge strength distribution near its edge, has been studied in the finite-dimensional case in [4]. The methods developed in this paper, in particular for the upperbound proofs, do not seem to us well adapted to the single edge dominated regime, which shall not been discussed further here.

B. Upper bound

The derivation of the upperbound in Eq. (11) will proceed by a decomposition of the sum over the walks $\omega \in \mathcal{M}_n$ according to the size $|\sigma(\omega)|$ of their support, i.e. the number of edges of the tree visited at least once by ω (recall the definitions given in Sec. III A). Indeed, the weight of a walk can be upperbounded as follows:

$$\pi(\omega) = \prod_{e\in\sigma(\omega)} \mathbb{E}[J^{2n_e(\omega)}] \leq \prod_{e\in\sigma(\omega)} \mathbb{E}[J^2] = \mathbb{E}[J^2]^{|\sigma(\omega)|} \ , \tag{33}$$

where we have used the facts that $|J| \leq 1$ and that (by definition of the support) $n_e(\omega) \geq 1$ for $e \in \sigma(\omega)$. By our assumption (OD1) the variance $\mathbb{E}[J^2]$ is strictly smaller than 1, hence the weight of a walk is exponentially small in the size of its support. The strategy of the proof corresponds then to choose in an optimal way a (n-dependent) threshold on the support's size, use the bound above for walks with larger supports, and upperbound the number of walks with smaller supports.

Let $r_{\rm m}(n)$ be an increasing (integer) function of n to be fixed later, such that $\lim r_{\rm m}(n) = \infty$ and $\lim \frac{r_{\rm m}(n)}{n} = 0$ as $n \to \infty$. We define $\mathcal{M}_n^{\leq} \subset \mathcal{M}_n$ as the set of walks ω whose support has size at most $r_{\rm m}(n)$. Using the obvious fact that $\pi(\omega) \leq 1$ for any walk, we obtain

$$c_n \leq \mathbb{E}[J^2]^{r_{\rm m}(n)} |\mathcal{M}_n \setminus \mathcal{M}_n^{\leq}| + |\mathcal{M}_n^{\leq}| \leq \mathbb{E}[J^2]^{r_{\rm m}(n)} |\mathcal{M}_n| + |\mathcal{M}_n^{\leq}| \ . \tag{34}$$

According to the discussion of Sec. III A, the number of walks in $\mathcal{M}_n^{\leq}$ can be written as

$$|\mathcal{M}_n^{\leq}| = \sum_{r=1}^{r_{\rm m}(n)} \sum_{\sigma\in\mathcal{S}^r} \sum_{\widehat{\sigma}\in\widehat{\mathcal{S}}_n^r(\sigma)} \eta(\widehat{\sigma}) \ , \tag{35}$$

where $\mathcal{S}^r$ is the set of supports of size r, $\widehat{\mathcal{S}}_n^r(\sigma)$ the set of skeletons of length n based on a support σ of size r, and $\eta(\widehat{\sigma})$ the number of walks compatible with such a skeleton. The rest of the proof will rely on the following statement about the maximal value $\kappa(n)$ of the combinatorial factor $\eta(\widehat{\sigma})$:

Let $\kappa(n) = \max\{\eta(\widehat{\sigma}) | r \in [1, n/2], \sigma \in \mathcal{S}^r, \widehat{\sigma} \in \widehat{\mathcal{S}}_n^r(\sigma)\}$. Then for all $y > 0$ and $\alpha < (\pi \ln k)^2/2$, it holds for n large enough that:

$$(2\sqrt{k})^n e^{-\frac{n}{(\ln n)^{2-y}}} \le \kappa(n) \le (2\sqrt{k})^n e^{-\alpha \frac{n}{(\ln n)^2}} \ . \tag{36}$$

It should be emphasized that the total number of walks of length n is, at the leading exponential order, $(2\sqrt{k})^n$, so that the previous statement shows that an unique skeleton $\widehat{\sigma}$ with very large combinatorial factor $\eta(\widehat{\sigma})$ contributes to the leading exponential order to the sum over all walks. In the next subsection we explain how to obtain these bounds on $\kappa(n)$; unfortunately we were not able to derive the upperbound in Eq. (36) in a fully rigorous way, yet it is the result of an analytical computation that we also checked numerically with high precision. Hence we continue here the derivation assuming that Eq. (36) holds true.

Going back to Eq. (35) and using the definition of $\kappa(n)$ yields

$$|\mathcal{M}_n^{\le}| \le \kappa(n) \sum_{r=1}^{r_{\mathrm{m}}(n)} \sum_{\sigma \in \mathcal{S}^r} |\widehat{\mathcal{S}}_n^r(\sigma)| \ . \tag{37}$$

It is easy to see that $|\widehat{\mathcal{S}}_n^r(\sigma)| = \binom{n/2-1}{r-1}$ independently of σ: this is the number of ways to choose r positive integers (the $\{n_e\}_{e\in\sigma}$) which sum to $n/2$. The numbers $|\mathcal{S}^r|$ of supports of size r are clearly increasing with r. Moreover $r_{\mathrm{m}}(n) < n/2$ for n large enough, hence the terms of the sum in Eq. (37) can be upperbounded by their value in $r_{\mathrm{m}}(n)$:

$$|\mathcal{M}_n^{\le}| \le \kappa(n) r_{\mathrm{m}}(n) \binom{n/2-1}{r_{\mathrm{m}}(n)-1} |\mathcal{S}^{r_{\mathrm{m}}(n)}| \ . \tag{38}$$

The numbers $|\mathcal{S}^r|$ of supports of size r can be enumerated via their generating function $F(x) = \sum_r |\mathcal{S}^r| x^r$. A support of size r being a subtree of $\mathbb{T}_k$ of r edges, it is composed of a number $j \in [0, k+1]$ of edges around the root, along with j subtrees of the copy of $\widetilde{\mathbb{T}}_k$ rooted at the neighbors of the root. This implies that $F(x) = (1 + x\widetilde{F}(x))^{k+1}$, where $\widetilde{F}$ is the equivalent of F for $\widetilde{\mathbb{T}}_k$, solution of $\widetilde{F}(x) = (1 + x\widetilde{F}(x))^k$. The asymptotic behavior of $|\mathcal{S}^r|$ for large r can be inferred from the analysis of the singularities of the generating function (see Theorem VII.3 in [51]). An elementary study of the reciprocal function $x(\widetilde{F}) = (\widetilde{F}^{1/k} - 1)/\widetilde{F}$ shows that it has a maximum equal to $1/\gamma_k$, with $\gamma_k = \frac{k^k}{(k-1)^{k-1}}$, hence $\widetilde{F}(x)$ has a square-root singularity in $x = 1/\gamma_k$. As $F(x)$ is known explicitly in terms of $\widetilde{F}$ one obtains easily

$$F(x) = A_k - B_k \sqrt{1 - \gamma_k x} + O(1 - \gamma_k x) \quad \text{as} \quad x \to 1/\gamma_k \ , \tag{39}$$

where A_k and B_k are two positive constants whose precise expression will not be useful in the following. This singularity is then translated in terms of the $|\mathcal{S}^r|$ as [51]:

$$|\mathcal{S}^r| = \frac{B_k}{2\sqrt{\pi r^3}} \gamma_k^r (1 + O(r^{-1})) \ . \tag{40}$$

We now choose an arbitrary $x > 0$, and set $r_{\mathrm{m}}(n) = \lfloor \frac{n}{(\ln n)^{2+x}} \rfloor$. With this choice for the maximal size of the supports in $\mathcal{M}_n^{\le}$, one has

$$\frac{1}{n} \ln \left[r_{\mathrm{m}}(n) \binom{n/2-1}{r_{\mathrm{m}}(n)-1} |\mathcal{S}^{r_{\mathrm{m}}(n)}| \right] = O\left(\frac{\ln \ln n}{(\ln n)^{2+x}} \right) \ . \tag{41}$$

Hence the two leading orders of the inequality in (38) arises from $\kappa(n)$. Thanks to the freedom of choice of α in Eq. (36) we conclude that for all $\beta < (\pi \ln k)^2/2$ and n large enough:

$$|\mathcal{M}_n^{\le}| \le (2\sqrt{k})^n e^{-\beta \frac{n}{(\ln n)^2}} \ . \tag{42}$$

Dividing Eq. (34) by $c_n^0 = |\mathcal{M}_n|$, the moment of the pure-case, one obtains for n large enough,

$$\frac{c_n}{c_n^0} \le e^{\lfloor \frac{n}{(\ln n)^{2+x}} \rfloor \ln(\mathbb{E}[J^2])} + e^{-\beta \frac{n}{(\ln n)^2}} \ , \tag{43}$$

using the fact that the sub-exponential corrections to c_n^0 are negligible here (recall that $\frac{1}{n}\ln(c_n^0) = \ln(2\sqrt{k}) + O(\frac{\ln n}{n})$). In this last inequality the first term is the dominating one, and leads to

$$\frac{1}{n}\ln(c_n) \le \ln(2\sqrt{k}) - \frac{1}{(\ln n)^{2+x}}\ln(\mathbb{E}[J^2]^{-1}) \ . \tag{44}$$

As x is constrained to be positive the (positive) constant $\ln(\mathbb{E}[J^2]^{-1})$ is irrelevant for n large enough; this yields the upperbound stated in Eq. (11).

Let us finally comment on the possible improvement of this upperbound. With a better estimate of $|\mathcal{M}_n \setminus \mathcal{M}_n^{\le}|$ one should try to make the threshold function $r_{\rm m}(n)$ grow faster with n. The limiting factor would then be the allowed range of β in the second term of Eq. (43), $\beta < (\pi \ln k)^2/2$; this would actually be the best one can hope for, as a greater value of β would contradict the lowerbound in Eq. (11).

C. Combinatorial factor

We shall now present our arguments in favor of the right inequality in (36). We recall that the expression of $\eta(\widehat{\sigma})$, i.e. the number of walks compatible with the skeleton $\widehat{\sigma}$, is given by the product of the multinomial coefficients in Eq. (22). The maximal value of η can thus be written as

$$\kappa(n) = \max_{r\in[1,n/2]}\max_{\sigma\in\mathcal{S}^r}\max_{\substack{\{n_e>0\}_{e\in\sigma}\\ 2\sum n_e = n}}\left[\binom{n_1+n_2+\cdots+n_{k+1}}{n_1,n_2,\ldots,n_{k+1}}\prod_{v\in\sigma\setminus 0}\binom{n_v-1+n_{v_1}+\cdots+n_{v_k}}{n_v-1,n_{v_1},\ldots,n_{v_k}}\right] , \tag{45}$$

where we follow the labelling of $\mathbb{T}_k$ introduced in Sec. III C, giving to an edge the index of its endpoint vertex most distant from the root. We also set by convention $n_e = 0$ if $e \notin \sigma$.

An upperbound on $\kappa(n)$ is obtained by relaxing the constraint on the possible values of the variables $\{n_e\}$. We shall thus extend their domain to non-negative reals, denoted x_e to avoid confusion, using the natural extension $\Gamma(x+1) = x!$ from integer x to arbitrary reals. We also let the set of edges with positive values of x_e to be an arbitrary subset of $\mathcal{E}_p$, which contains the edges with endpoint at distance at most p from the root. For the ease of notation we introduce $\varkappa(n) = \ln\kappa(n)$, which is thus upperbounded as

$$\begin{aligned}\varkappa(n) \le \max_{p\in[1,n/2]}\sup_{\substack{\{x_e\ge 1\}_{e\in\mathcal{E}_p}\\ 2\sum x_e = n}}\Bigg[&\ln\Gamma(1+x_1+x_2+\cdots+x_{k+1}) - \ln\Gamma(1+x_1) - \cdots - \ln\Gamma(1+x_{k+1}) \\ &+ \sum_{v\in\mathcal{E}_{p-1}\setminus 0}[\ln\Gamma(x_v+x_{v_1}+\cdots+x_{v_k}) - \ln\Gamma(x_v) - \ln\Gamma(1+x_{v_1}) - \cdots - \ln\Gamma(1+x_{v_k})]\Bigg] .\end{aligned} \tag{46}$$

We expect this upperbound to become tight at the leading order when $n\to\infty$: the integer values of n_e achieving the maximum in Eq. (45) will become large as well, hence relaxing them to be real numbers should have a minor effect.

For a given p the supremum is over a smooth function of the $|\mathcal{E}_p|$ reals x_e; assuming that the supremum is reached in the interior of its domain, one has to look for the critical points of the function. There is always one radially symmetric critical point, i.e. where x_e depends on e only through the depth i of the edge; we denote $\widehat{x}_i$ the common value of x_e for all edges at depth i. We assume that this is the global maximum of the function. This yields

$$\varkappa(n) \le \varkappa_*(n) = \max_{p\in[1,n/2]}\sup_{\substack{\widehat{x}_1,\ldots,\widehat{x}_p\ge 0\\ 2\sum_{i=1}^p v_i\widehat{x}_i = n}} K_p(\widehat{x}_1,\ldots,\widehat{x}_p) , \tag{47}$$

where we defined the radially symmetric function K_p as

$$K_p(\widehat{x}_1,\ldots,\widehat{x}_p) = \sum_{i=0}^{p-1} v_i[\ln\Gamma(\widehat{x}_i + k\widehat{x}_{i+1}) - \ln\Gamma(\widehat{x}_i) - k\ln\Gamma(1+\widehat{x}_{i+1})] . \tag{48}$$

For $i\ge 1$ we denoted $v_i = (k+1)k^{i-1}$ the number of edges of depth i, and we set by convention $v_0 = 1$ and $\widehat{x}_0 = \widehat{x}_1 + 1$. For a given p the critical point of K_p is achieved at the solution of:

$$\psi(\widehat{x}_i + k\widehat{x}_{i+1}) - \psi(\widehat{x}_i) + \psi(\widehat{x}_{i-1} + k\widehat{x}_i) - \psi(\widehat{x}_i + 1) = \lambda \quad \text{for} \ \ i\in[1,p] \text{ with } \ \widehat{x}_0 = 1+\widehat{x}_1 \ , \ \ \widehat{x}_{p+1} = 0 \ , \tag{49}$$

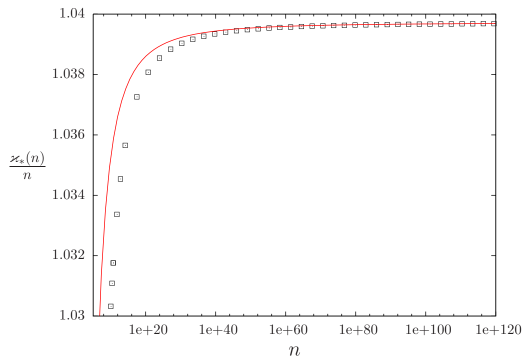


FIG. 6: Plot of the function $\varkappa_*(n)$ defined in Eq. (47), the symbols are the result of the numerical evaluation of $\varkappa_*(n)/n$, the line is the conjecture (51). These results are for $k = 2$.

where λ is a Lagrange multiplier that has to be fixed to the value enforcing the constraint $\sum_{i=1}^{p} v_i \widehat{x}_i = n/2$, and $\psi \equiv \frac{\mathrm{d} \ln \Gamma}{\mathrm{d}x}$ is the digamma function. Note that $\varkappa_*$ is defined by extension for real values of n, and that λ plays the role of a parameter conjugated to n. The previous equation can be rewritten as a three terms recurrence on $\widehat{x}_i$:

$$\widehat{x}_{i+1} = \frac{1}{k}\left[-\widehat{x}_i + \psi^{-1}(\lambda + \psi(\widehat{x}_i) + \psi(\widehat{x}_i + 1) - \psi(\widehat{x}_{i-1} + k\widehat{x}_i))\right] \ . \tag{50}$$

We solved numerically the optimization problem defined in Eq. (47); the plot in Fig. 6 display the values of $\varkappa_*(n)/n$ we found in this way, as well as the good agreement with our conjecture on its asymptotic behavior:

$$\frac{\varkappa_*(n)}{n} = \ln(2\sqrt{k}) - \frac{(\pi \ln k)^2}{2(\ln n)^2} + o\left(\frac{1}{(\ln n)^2}\right) \ , \tag{51}$$

which implies the upperbound stated in Eq. (36). We turn now to the computation that led us to the formula in Eq. (51). In the large n limit most of the non-zero $\widehat{x}_i$ can be expected to be large (this is also confirmed by the numerical computation of $\varkappa_*$). In consequence we can simplify the recurrence relation (50) using the first term of the asymptotic expansion of the digamma function, $\psi(x) = \ln(x) + O(1/x)$. The resulting relation is more compactly written in terms of $\alpha_i = \frac{\widehat{x}_i}{\widehat{x}_{i-1}}$, which is found to obey

$$\alpha_{i+1} = f(\alpha_i, \lambda) \ , \quad \text{with} \ \ f(\alpha, \lambda) = \frac{1}{k}\left(-1 + \frac{e^{\lambda}}{k + \frac{1}{\alpha}}\right) \ . \tag{52}$$

At the order of our approximation the initial condition $\alpha_1 = \frac{\widehat{x}_1}{1+\widehat{x}_1}$ is equal to $\alpha_1 = 1$. The fixed-point equation $\alpha = f(\alpha, \lambda)$ undergoes a bifurcation transition at a critical value $\lambda_{\mathrm{c}} = \ln(4k)$ (see left panel in Fig. 7). For $\lambda > \lambda_{\mathrm{c}}$ there are two solutions which coalesce in $\alpha_* = \frac{1}{k}$ at λ_{c}, and disappear for $\lambda < \lambda_{\mathrm{c}}$. This bifurcation has the following consequences on the solution of Eq. (52) with initial condition $\alpha_1 = 1$, that for clarity we denote $\alpha_i(\lambda)$ to emphasize its dependency on the parameter λ:

- right at the transition the sequence decays towards its fixed-point as $1/i$, more precisely

$$\alpha_i(\lambda_{\mathrm{c}}) = \frac{1}{k} + \frac{2}{k}\frac{1}{i} + o\left(\frac{1}{i}\right) \ . \tag{53}$$

- when $\lambda \to \lambda_{\mathrm{c}}^-$ a plateau develops around the avoided fixed-point $\alpha_* = 1/k$, with a diverging length of order $(\lambda_{\mathrm{c}} - \lambda)^{-1/2}$, see the right panel of Fig. 7. One can establish the following scaling behavior by zooming around the plateau:

$$\lim_{\lambda \to \lambda_{\mathrm{c}}^-} \left[\frac{1}{\sqrt{\lambda_{\mathrm{c}} - \lambda}}\left(\alpha_{i=\frac{\theta}{\sqrt{\lambda_{\mathrm{c}}-\lambda}}}(\lambda) - \frac{1}{k}\right)\right] = -\frac{2}{k}\tan\left(\theta - \frac{\pi}{2}\right) \quad \forall \ \ \theta \in]0, \pi[\ . \tag{54}$$

This in particular shows that the length of the plateau is $\pi/\sqrt{\lambda_{\mathrm{c}} - \lambda}$ at the leading order. We shall use this scaling behavior under the less precise but more readable form:

$$\alpha_i \approx \frac{1}{k} - \frac{2}{k}\sqrt{\lambda_{\mathrm{c}} - \lambda}\tan\left(i\sqrt{\lambda_{\mathrm{c}} - \lambda} - \frac{\pi}{2}\right) \ , \tag{55}$$

where it is understood that i is in the scaling regime, i.e. of the form $\theta/\sqrt{\lambda_{\mathrm{c}} - \lambda}$ with $\theta \in]0, \pi[$.

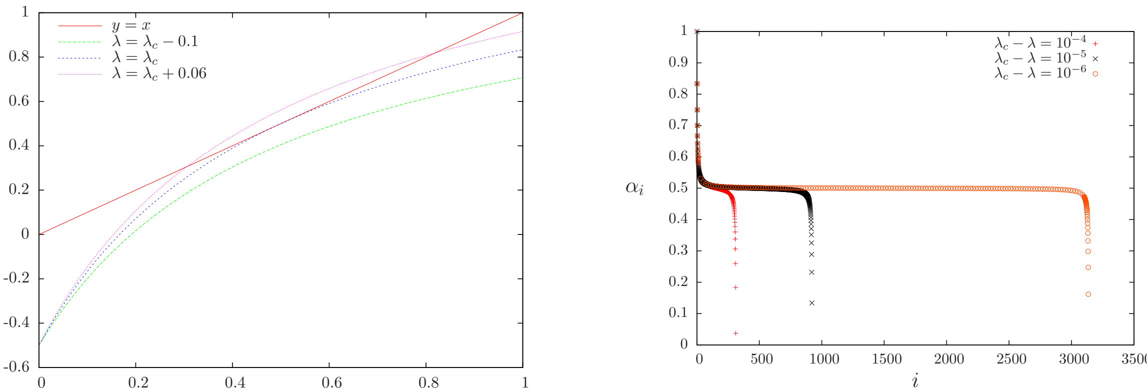


FIG. 7: Left panel: $f(x,\lambda)$ for different λ above, at and below λ_c. Right panel: the sequences α_i solution of Eq. (52) for values of λ slightly smaller than λ_c. All data are for $k = 2$.

Note that there is a matching condition between these two regimes: if one considers a value of i very large but finite with respect to $(\lambda_{\rm c} - \lambda)^{-1/2}$, the same behavior of $\alpha_i(\lambda_{\rm c})$ is obtained by taking $i \to \infty$ in Eq. (53) or $\lambda \to \lambda_{\rm c}$ with i fixed in Eq. (55). The above behaviors are generic for all recursion of the forms $\alpha_{i+1} = f(\alpha_i, \lambda)$ that encounters a bifurcation transition, and the quantitative statements only depend on the value of the partial derivatives $\partial_\lambda f$ and $\partial^2_{\alpha,\alpha} f$ of f in $(\alpha_*, \lambda_{\rm c})$. It turns out in addition that for the specific function f of Eq. (52) the critical sequence can be computed exactly, $\alpha_i(\lambda_{\rm c}) = ((k+1) + (k-1)i)/(k(3-k) + k(k-1)i)$.

Let us now investigate the consequence of these behaviors of $\alpha_i(\lambda)$ on the properties of the $\widehat{x}_i(\lambda)$. The leading order of the recurrence relation (50) only constrains the ratio α_i between successive terms in the sequence $\{\widehat{x}_i\}$, in consequence the initial condition $\widehat{x}_1$ is at this point a free parameter that we shall fix afterwards. By the definition of α_i we have

$$\widehat{x}_i(\lambda) = \widehat{x}_1 \left(\frac{1}{k}\right)^{i-1} \prod_{j=2}^{i} (k\alpha_j(\lambda)) \ , \tag{56}$$

where we have factorized the plateau value $1/k$ of the α_i. Consider first the large i regime at the transition, i.e. for $\lambda = \lambda_{\rm c}$. As a consequence of (53) one has $\sum_{j=2}^{i} \ln(k\alpha_j(\lambda_{\rm c})) = 2\ln(i) + \ln(C_k) + O(1/i)$, hence

$$\widehat{x}_i(\lambda_{\rm c}) = \widehat{x}_1 \left(\frac{1}{k}\right)^{i-1} i^2 C_k (1 + O(1/i)) \ , \tag{57}$$

where C_k is a positive k-dependent constant. Actually one can compute C_k explicitly thanks to the exact expression of $\alpha_i(\lambda_{\rm c})$, and find $C_k = (k-1)^2/(2k(k+1))$. Let us now turn to the scaling regime $\lambda \to \lambda_{\rm c}^-$ where Eq. (55) is valid. Consider two indices $i_0 < i$ of the form $i_0 = \theta_0/\sqrt{\lambda_{\rm c} - \lambda}$ and $i = \theta/\sqrt{\lambda_{\rm c} - \lambda}$. One has

$$\begin{aligned} \widehat{x}_i(\lambda) &= \widehat{x}_{i_0}(\lambda) \left(\frac{1}{k}\right)^{i-i_0} \exp\left[\sum_{j=i_0+1}^{i} \ln(k\alpha_j(\lambda))\right] && (58) \\ &\approx \widehat{x}_{i_0}(\lambda) \left(\frac{1}{k}\right)^{i-i_0} \exp\left[-2\sqrt{\lambda_{\rm c} - \lambda} \sum_{j=i_0+1}^{i} \tan\left(j\sqrt{\lambda_{\rm c} - \lambda} - \frac{\pi}{2}\right)\right] \ , && (59) \end{aligned}$$

where we have used Eq. (55) in the second line. Now in the limit $\lambda \to \lambda_{\rm c}$ the term in square brackets is the Riemann discretization of the integral of $\tan(\theta - \pi/2)$, hence

$$\widehat{x}_i(\lambda) \approx \widehat{x}_{i_0}(\lambda) k^{i_0 - 1} \frac{1}{\sin^2(\theta_0)} \left(\frac{1}{k}\right)^{i-1} \sin^2(\theta) \ . \tag{60}$$

Finally we invoke a matching argument as explained above : we take i_0 large but finite with respect to $(\lambda - \lambda_{\rm c})^{-1/2}$, so that $\theta_0 \to 0$, and we replace the value of $\widehat{x}_{i_0}$ by its value computed at $\lambda_{\rm c}$ in Eq. (57). This yields

$$\widehat{x}_i(\lambda) \approx \frac{\widehat{x}_1 C_k}{\lambda_{\rm c} - \lambda} \left(\frac{1}{k}\right)^{i-1} \sin^2(i\sqrt{\lambda_{\rm c} - \lambda}) \ , \tag{61}$$

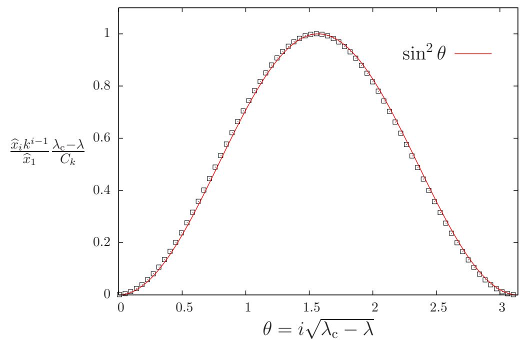


FIG. 8: Verification of Eq. (61) for $k = 2$, $\lambda = \lambda_c - 4.10^{-5}$. The solid line is the analytical prediction, the symbols are the $\widehat{x}_i$ obtained from the resolution of the optimization problem of Eq. (47), and are indistinguishable on this scale from their approximate computation in terms of the α_i. Note that there is no fitting parameter in this figure.

valid in the scaling regime $i = \theta/\sqrt{\lambda_c - \lambda}$ with $\theta \in]0, \pi[$. In Fig. 8 we present a numerical confirmation of this prediction.

As can be seen on the right panel of Fig. 7, once the the sequence α_i exits the scaling regime around the plateau a finite number of additional iterations brings it to negative values. Let us call $p(\lambda)$ the largest value of i such that $\alpha_i(\lambda)$ is positive. From the observation above, $p(\lambda) = \pi/\sqrt{\lambda_c - \lambda}$ at the leading order when $\lambda \to \lambda_c^-$. Let us now discuss the scaling of $\widehat{x}_1$. For the computation above to be consistent with the replacement of $\psi(x)$ by $\ln(x)$ it was based upon, one should have $\widehat{x}_i(\lambda)$ large for most of the i's between 1 and $p(\lambda)$. On the other hand one should impose in an approximated way the boundary condition $\widehat{x}_{p(\lambda)+1} = 0$ of the exact system (49). In consequence we shall choose $\widehat{x}_1$ such that $\widehat{x}_i$ is of order one for i at the end of the scaling regime ($i = \pi^-/\sqrt{\lambda_c - \lambda}$), hence at the leading order (within multiplicative constant and sub-exponential corrections in $(\lambda_c - \lambda)^{-1/2}$)

$$\widehat{x}_1(\lambda) \approx k^{\frac{\pi}{\sqrt{\lambda_c-\lambda}}} \ , \quad \text{i.e.} \qquad \lambda_c - \lambda \sim \frac{(\pi \ln k)^2}{(\ln \widehat{x}_1)^2} \ . \tag{62}$$

This ends our determination of the asymptotic form of the optimal value of the sequence $\{\widehat{x}_i\}_{i=1}^{p(\lambda)}$ as a function of λ; from it we shall now compute the corresponding values of n and $\varkappa_*$. The former reads

$$n(\lambda) = 2(k+1)\sum_{i=1}^{p(\lambda)} k^i \widehat{x}_i(\lambda) \approx 2\frac{k+1}{k}\frac{C_k \widehat{x}_1(\lambda)}{(\lambda_c-\lambda)^{3/2}}\int_0^\pi \mathrm{d}\theta \ \sin^2(\theta) = \pi\frac{k+1}{k}\frac{C_k\widehat{x}_1(\lambda)}{(\lambda_c-\lambda)^{3/2}} \ , \tag{63}$$

where we replaced the sum over i by an integral over θ using the form (61) of the $\widehat{x}_i$ in the scaling regime. The leading behavior of n is thus dictated by the one of $\widehat{x}_1(\lambda)$, and the relation (62) can thus be re-expressed as

$$\lambda_c - \lambda \sim \frac{(\pi \ln k)^2}{(\ln n)^2} \ . \tag{64}$$

To compute the value of $\varkappa_*$ achieved in the large n (or equivalently $\lambda \to \lambda_c^-$) limit we use the asymptotic expansion $\ln\Gamma(x) = x\ln x - x + O(\ln x)$ (consistently with our approximation of $\psi(x)$), the term in square brackets in (48) reading

$$\widehat{x}_i[\ln(4k) + \beta_{i+1}\ln(2k) + (2+\beta_{i+1})\ln\left(1+\frac{\beta_{i+1}}{2}\right) - (1+\beta_{i+1})\ln(1+\beta_{i+1})] + O(\ln\widehat{x}_i) \ , \tag{65}$$

where we defined β_i by $\alpha_i = \frac{1}{k}(1+\beta_i)$. Once inserted in the sum over i the first term leads to a factor proportional to n,

$$\varkappa_*(n) \approx n\ln(2\sqrt{k}) + \sum_i v_i\widehat{x}_i[\beta_{i+1}\ln(2k) + (2+\beta_{i+1})\ln\left(1+\frac{\beta_{i+1}}{2}\right) - (1+\beta_{i+1})\ln(1+\beta_{i+1})] \ . \tag{66}$$

In the scaling regime $\beta_i = -2\sqrt{\lambda_c-\lambda}\tan(i\sqrt{\lambda_c-\lambda}-\frac{\pi}{2})$ is small, we thus expand the previous expression to second order in β_i and trade the sum with an integral to obtain

$$\varkappa_*(n) \approx n\ln(2\sqrt{k}) + \frac{k+1}{k}\frac{\widehat{x}_1(\lambda)C_k}{(\lambda_c-\lambda)^{3/2}}\int_0^\pi \mathrm{d}\theta\,\sin^2(\theta)\left[-2\ln(2k)\sqrt{\lambda_c-\lambda}\tan\left(\theta-\frac{\pi}{2}\right) - (\lambda_c-\lambda)^2\tan^2\left(\theta-\frac{\pi}{2}\right)\right] \ . \tag{67}$$

The integral of the first term vanishes for symmetry reasons, the second one is easily evaluated, and by comparison with the expression of n given in (63) one obtains

$$\frac{\varkappa_*(n)}{n} \approx \ln(2\sqrt{k}) - \frac{1}{2}(\lambda_c - \lambda) \ . \tag{68}$$

With the expression of $\lambda_c - \lambda$ as a function of n given in (64), this completes the derivation of (51).

Finally the lowerbound in (36) can easily be justified by contradiction: suppose there exists $x > 0$ such that for arbitrarily large n, $\varkappa(n) \leq (2\sqrt{k})^n e^{-\frac{n}{(\ln n)^{2-x}}}$. Then one could repeat the derivation of Sec. IV B, choosing $r_m(n) = \lfloor \frac{n}{(\ln n)^{2-\frac{x}{2}}} \rfloor$, and obtain instead of (44) the upperbound

$$\frac{1}{n}\ln(c_n) \leq \ln(2\sqrt{k}) - \frac{1}{(\ln n)^{2-\frac{x}{2}}} \ln(\mathbb{E}[J^2]^{-1}) \ . \tag{69}$$

This would violate the lowerbound in (11), hence the contradiction.

V. PROOFS FOR THE DIAGONAL DISORDER MODEL

This section contains the proofs of the bounds stated in Eq. (12) for the moments d_n of the diagonal disorder case, i.e. for edge couplings J_{ij} deterministically equal to 1 and vertex random variables V_i drawn in $[0, W]$ with a probability distribution satisfying the assumptions (D1-D2) of Sec. II A. The proofs are similar to the off-diagonal case, we mostly precise the additional technicalities that arises here.

A. Lower bound

As in the off-diagonal case the lower bound on d_n follows by considering only the walks restricted to a neighborhood of depth h around the root of the tree, and by conditioning on the event that the disorder is large in this neighborhood.

We recall that $d_n = \sum_{\omega \in \mathcal{W}_n} \pi(\omega)$, where $\mathcal{W}_n$ denotes now the set of closed walks on $\mathbb{T}_k$ of length n with self-bond steps allowed, and the weight of a walk is $\pi(\omega) = \mathbb{E}[\prod_{v \in \mathcal{V}} V_v^{s_v(\omega)}]$, with $s_v(\omega)$ the number of self-bond steps taken by the walk around the vertex v of the tree. These weights being positive we can lowerbound d_n as

$$d_n \geq \sum_{\omega \in \widetilde{\mathcal{W}}_{n,h}} \pi(\omega) \ , \tag{70}$$

where $\widetilde{\mathcal{W}}_{n,h}$ is the set of walks on $\widetilde{\mathbb{T}}_k$ that visit vertices at distance at most h from the root. We call $\mathcal{V}_h$ the set of such vertices, introduce a positive threshold $V_0 < W$, and define the event M as $\{V_v \geq V_0 \ \forall v \in \mathcal{V}_h\}$. With a reasoning similar to the one which yielded Eq. (27) we obtain

$$\pi(\omega) \geq \mathbb{P}[V \geq V_0]^{|\mathcal{V}_h|} \, V_0^{s_{\rm tot}(\omega)} \tag{71}$$

for all walks $\omega \in \widetilde{\mathcal{W}}_{n,h}$, where we defined $s_{\rm tot}(\omega) = \sum_v s_v(\omega)$ the total number of self-bond steps taken by the walk. Putting these two inequalities together one has

$$d_n \geq \mathbb{P}[V \geq V_0]^{|\mathcal{V}_h|} \sum_{\omega \in \widetilde{\mathcal{W}}_{n,h}} V_0^{s_{\rm tot}(\omega)} \ . \tag{72}$$

We now use the projection from a walk to its depth first introduced in Sec. III D. The self-bond steps are then associated to horizontal steps in the path, which is called a Motzkin path in this case. The sum over ω in the last equation is thus equal to the weighted sum over Motzkin path of length n, where ascending steps have weight k (the number of branches the walk can choose from in a step away from the root) and horizontal (resp. descending) steps have weight V_0 (resp. 1). This quantity $m_{n,h}$ is computed by combinatorial techniques in App. E. The number of vertices at depth at most h is $|\mathcal{V}_h| = \frac{k^{h+1}}{k-1}$, hence

$$\frac{1}{n}\ln d_n \geq \frac{1}{n}\ln m_{n,h} + \frac{k^{h+1}}{n(k-1)} \ln \mathbb{P}[V \geq V_0] \ . \tag{73}$$

If we let $n \to \infty$ with h and V_0 fixed, the asymptotic properties of $m_{n,h}$ proved in App. E (see Eq. (E8)) yields

$$\liminf_{n\to\infty} \frac{1}{n} \ln d_n \geq \ln\left(V_0 + 2\sqrt{k}\cos\left(\frac{\pi}{h+2}\right)\right) \ . \tag{74}$$

Letting finally $h \to \infty$ and $V_0 \to W$ and noting that $d_n \leq d_n^0$, the shifted pure model with $J_{ij} = 1$ and $V_i = W$ for which $\lim \frac{1}{n} \ln d_n^0 = \ln(2\sqrt{k} + W)$ allows to conclude $\lim \frac{1}{n} \ln d_n = \ln(2\sqrt{k} + W)$. The upper limit of the support is thus $2\sqrt{k} + W$ in the diagonal disorder case, in agreement with the ergodicity arguments sketched in Sec. II B.

We now consider the limit $n \to \infty$, taking the parameters $h \to \infty$ and $V_0 \to W$ in an n-dependent way. For the same reasons as in the off-diagonal case the optimal scaling of the parameters is $h = \lfloor \alpha \ln n \rfloor$ with $\alpha < 1/\ln k$ and $V_0 = W - \frac{1}{(\ln n)^{2+y}}$ where $y > 0$ is an arbitrary positive constant. Using the asymptotic expansion (E9) for $m_{n,h}$, the r.h.s. of (73) becomes

$$\ln(2\sqrt{k} + W) - \frac{\pi^2\sqrt{k}}{\alpha^2(2\sqrt{k}+W)} \frac{1}{(\ln n)^2} + \frac{k}{k-1} \frac{1}{n^{1-\alpha \ln k}} \ln \mathbb{P}\left[V \geq W - \frac{1}{(\ln n)^{2+y}}\right] + O\left(\frac{1}{(\ln n)^{2+y}}\right) \ . \tag{75}$$

The assumption (D2) on the random variable V ensures that the third term is negligible for any $\alpha < 1/\ln k$ and $y > 0$, which concludes the proof of the lower bound in Eq. (12).

B. Upper bound

We present now the proof of the upper bound in Eq. (12). Let us first recall the definition of the moment of the shifted pure case ($J_{ij} = 1$, $V_i = W$) in terms of the weights $\pi^0(\omega)$ of the walks:

$$d_n^0 = \sum_{\omega \in \mathcal{W}_n} \pi^0(\omega) \ , \qquad \pi^0(\omega) = \prod_{v \in \mathcal{V}} W^{s_v(\omega)} = W^{s_{\rm tot}(\omega)} \ , \tag{76}$$

d_n^0 being equal to $(2\sqrt{k} + W)^n$ within polynomial corrections. Consider now the effect of the disorder on the weight of a walk:

$$\pi(\omega) = \prod_{v \in \mathcal{V}} \mathbb{E}[V^{s_v(\omega)}] = \pi^0(\omega) \prod_{v \in \mathcal{V}} \mathbb{E}\left[\left(\frac{V}{W}\right)^{s_v(\omega)}\right] \leq \pi^0(\omega) \left(\mathbb{E}\left[\frac{V}{W}\right]\right)^{\tau(\omega)} \ , \tag{77}$$

where $\tau(\omega) = |\{v \in \mathcal{V} | s_v(\omega) \geq 1\}$ is the size of the self-support of the walk, as defined in Sec. III A. By our assumption (D1) $\mathbb{E}[V/W]$ is strictly smaller than 1, hence the weight of a walk is exponentially suppressed in the size of its self-support. The proof will thus be based on partitioning the set of walks in $\mathcal{W}_n$ between those with a large self-support, for which we shall use the bound above, and on bounding the contribution of the walks with small self-supports.

Let us introduce a growing integer function $t_{\rm m}(n)$, with $t_{\rm m}(n) \to \infty$ and $\frac{t_{\rm m}(n)}{n} \to 0$ as $n \to \infty$, and denote $\mathcal{W}_n^{\leq}$ the subset of $\mathcal{W}_n$ which contains the walks whose self-support contains less than $t_{\rm m}(n)$ vertices, $\mathcal{W}_n^{\leq} = \{\omega \in \mathcal{W}_n | \tau(\omega) \leq t_{\rm m}(n)\}$. Then

$$d_n \leq \left(\mathbb{E}\left[\frac{V}{W}\right]\right)^{t_{\rm m}(n)} \sum_{\omega \in \mathcal{W}_n \setminus \mathcal{W}_n^{\leq}} \pi^0(\omega) + \sum_{\omega \in \mathcal{W}_n^{\leq}} \pi^0(\omega) \leq \left(\mathbb{E}\left[\frac{V}{W}\right]\right)^{t_{\rm m}(n)} d_n^0 + \sum_{\omega \in \mathcal{W}_n^{\leq}} \pi^0(\omega) \ , \tag{78}$$

where we have used the obvious facts that $\pi(\omega) \leq \pi^0(\omega)$ and $\pi^0(\omega) \geq 0$ for all ω.

We shall further partition $\mathcal{W}_n^{\leq}$ according to the size of the supports of the walks (i.e. the number of edges which are crossed at least once by the walk). Let us introduce another growing threshold function $r_{\rm m}(n)$ with $r_{\rm m}(n) \to \infty$ and $\frac{r_{\rm m}(n)}{n} \to 0$, and define $\mathcal{W}_n^1 = \{\omega \in \mathcal{W}_n^{\leq} || \sigma(\omega)| \leq r_{\rm m}(n)\}$ and $\mathcal{W}_n^2 = \{\omega \in \mathcal{W}_n^{\leq} || \sigma(\omega)| > r_{\rm m}(n)\}$. We can thus refine the bound above as

$$d_n \leq \left(\mathbb{E}\left[\frac{V}{W}\right]\right)^{t_{\rm m}(n)} d_n^0 + \sum_{\omega \in \mathcal{W}_n^1} \pi^0(\omega) + \sum_{\omega \in \mathcal{W}_n^2} \pi^0(\omega) \ . \tag{79}$$

We shall work out separately upperbounds for the contributions of $\mathcal{W}_n^1$ and $\mathcal{W}_n^2$. Let us start with the former, and introduce a larger set of walks $\mathcal{W}_n^{1'} = \{\omega \in \mathcal{W}_n || \sigma(\omega)| \leq r_{\rm m}(n)\}$, in which we have removed the constraint on the

size of the self-supports. By the positivity of the weights the sum over $\mathcal{W}_n^{1'}$ dominates the one on $\mathcal{W}_n^1$. The crucial point is now to realize that there is a $\binom{n}{s}$ to 1 mapping between each walk $\omega \in \mathcal{W}_n$ with $s_{\rm tot}(\omega) = s$ and the walk $\overline{\omega} \in \mathcal{M}_{n-s}$ obtained by removing all self-bond steps, that in consequence shares the same skeleton. This binomial coefficient counts the number of possible choices for the s times $0 \leq t_1 \leq \cdots \leq t_s \leq n-s$ at which a self-bond step is inserted in $\overline{\omega}$ to construct ω. For completeness we give in App. F a more formal proof of this fact based on the study of the combinatorial factors. We can thus write

$$\sum_{\omega \in \mathcal{W}_n^1} \pi^0(\omega) \leq \sum_{\omega \in \mathcal{W}_n^{1'}} \pi^0(\omega) = \sum_{s=0}^{n} \binom{n}{s} W^s |\mathcal{M}_{n-s}^{\leq}| \ , \tag{80}$$

where $\mathcal{M}_{n-s}^{\leq}$ is the set of walks of length $n-s$, without self-bond steps, studied in Sec. IV B. We shall thus exploit the bound on $|\mathcal{M}_n^{\leq}|$ derived in that section: we choose an arbitrary $x > 0$, $\beta < (\ln k)^2 \pi^2/2$ and set $r_{\rm m}(n) = \lfloor \frac{n}{(\ln n)^{2+x}} \rfloor$. From Eq. (42) we obtain the existence of an (even) n_0 such that, for $n - s \geq n_0$,

$$|\mathcal{M}_{n-s}^{\leq}| \leq (2\sqrt{k})^{n-s} e^{-\beta \frac{n-s}{(\ln(n-s))^2}} \leq (2\sqrt{k})^{n-s} e^{-\beta \frac{n-s}{(\ln n)^2}} \ , \tag{81}$$

this bound being trivially true for odd values of $n-s$ for which $|\mathcal{M}_{n-s}^{\leq}| = 0$. Inserting this bound in the inequality above yields

$$\begin{aligned}
\sum_{\omega \in \mathcal{W}_n^1} \pi^0(\omega) &\leq \sum_{s=0}^{n-n_0} \binom{n}{s} W^s (2\sqrt{k})^{n-s} e^{-\beta \frac{n-s}{(\ln n)^2}} + \sum_{s=n-n_0+1}^{n} \binom{n}{s} W^s |\mathcal{M}_{n-s}^{\leq}| && (82)\\
&\leq \sum_{s=0}^{n} \binom{n}{s} W^s (2\sqrt{k})^{n-s} e^{-\beta \frac{n-s}{(\ln n)^2}} + |\mathcal{M}_{n_0}^{\leq}| \sum_{s=0}^{n} \binom{n}{s} W^s && (83)\\
&= e^{-\beta \frac{n}{(\ln n)^2}} (2\sqrt{k} + W e^{\beta \frac{1}{(\ln n)^2}})^n + |\mathcal{M}_{n_0}^{\leq}| (1+W)^n \ . && (84)
\end{aligned}$$

From the first to the second line we have used the fact that $|\mathcal{M}_n^{\leq}|$ grows with (even) n and extended the range of the sums, all terms being non-negative. As n_0 is now fixed and $2\sqrt{k} > 1$ the second term is exponentially smaller than the first one in the large n limit. Moreover the first one can be expanded as

$$\begin{aligned}
e^{-\beta \frac{n}{(\ln n)^2}} (2\sqrt{k} + W e^{\beta \frac{1}{(\ln n)^2}})^n &= e^{-\beta \frac{n}{(\ln n)^2}} (2\sqrt{k} + W)^n \left(1 + \frac{W}{2\sqrt{k}+W} \left(e^{\beta \frac{1}{(\ln n)^2}} - 1 \right) \right)^n && (85)\\
&= (2\sqrt{k} + W)^n e^{-\beta \frac{2\sqrt{k}}{2\sqrt{k}+W} \frac{n}{(\ln n)^2} + O(n/(\ln n)^4)} \ . && (86)
\end{aligned}$$

We can thus conclude the study of the sum over $\mathcal{W}_n^1$: for all $\gamma < \frac{(\pi \ln k)^2}{2} \frac{2\sqrt{k}}{2\sqrt{k}+W}$ it holds for n large enough that

$$\frac{1}{d_n^0} \sum_{\omega \in W_n^1} \pi^0(\omega) \leq e^{-\gamma \frac{n}{(\ln n)^2}} \ . \tag{87}$$

Let us finally derive an upperbound on the contributions of the walks in the set $\mathcal{W}_n^2$ whose support contains more than $r_{\rm m}(n)$ edges while their self-support contains less than $t_{\rm m}(n)$ vertices. We write this sum as

$$\sum_{\omega \in \mathcal{W}_n^2} \pi^0(\omega) = \sum_{s=0}^{n-2r_{\rm m}(n)-2} W^s \sum_{r=r_{\rm m}(n)+1}^{n/2} \sum_{\overline{\omega} \in \mathcal{M}_{n-s}^r} \sum_{\substack{\mathcal{A} \subset \sigma(\overline{\omega}) \cup 0 \\ |\mathcal{A}| \leq t_{\rm m}(n)}} E(\overline{\omega}, s, \mathcal{A}) \ , \tag{88}$$

where $\mathcal{M}_n^r$ is the set of walks of length n without self-bond steps with a support of r edges, and $E(\overline{\omega}, s, \mathcal{A})$ is the number of walks $\omega \in \mathcal{W}_n$ that are obtained from $\overline{\omega} \in \mathcal{M}_{n-s}$ by adding to it s self-bond steps, in such a way that the self-support of ω is a given $\mathcal{A}$, a subset of the vertices visited by $\overline{\omega}$. A short combinatorial reasoning presented in App. F yields the following upperbound for this quantity

$$E(\overline{\omega}, s, \mathcal{A}) \leq \binom{n - 2|\sigma(\overline{\omega})| + (k+1)|\mathcal{A}|}{s} \ . \tag{89}$$

In all the terms of (88) we have $|\sigma(\overline{\omega})| \geq r_{\rm m}(n)+1$ and $|\mathcal{A}| \leq t_{\rm m}(n)$; moreover the number of terms in the sum over $\mathcal{A}$ can be coarsely upperbounded by $t_{\rm m}(n)\binom{n}{t_{\rm m}(n)}$. This being done one can relax the constraint over r and sum over all walks $\overline{\omega} \in \mathcal{M}_{n-s}$, which are no more numerous than $(2\sqrt{k})^{n-s}$. This yields

$$\sum_{\omega \in \mathcal{W}_n^2} \pi^0(\omega) \leq \sum_{s=0}^{n-2r_{\rm m}(n)-2} W^s (2\sqrt{k})^{n-s} t_{\rm m}(n) \binom{n}{t_{\rm m}(n)} \binom{n-2r_{\rm m}(n)-2+(k+1)t_{\rm m}(n)}{s} \ . \tag{90}$$

We choose now the threshold function on the self-support size as $t_{\rm m}(n) = \lfloor \frac{n}{(\ln n)^{2+2x}} \rfloor$, which is in consequence negligible with respect to $r_{\rm m}(n)$ as $n \to \infty$. In order to perform the sum over s we decompose the last binomial coefficient as

$$\binom{n-2r_{\rm m}(n)-2+(k+1)t_{\rm m}(n)}{s} = \binom{n-2r_{\rm m}(n)-2}{s} \prod_{i=1}^{(k+1)t_{\rm m}(n)} \frac{n-2r_{\rm m}(n)-2+i}{n-2r_{\rm m}(n)-2-s+i} \tag{91}$$

$$\leq \binom{n-2r_{\rm m}(n)-2}{s} \frac{n^{(k+1)t_{\rm m}(n)}}{((k+1)t_{\rm m}(n))!} \ . \tag{92}$$

This yields

$$\sum_{\omega \in \mathcal{W}_n^2} \pi^0(\omega) \leq t_{\rm m}(n) \binom{n}{t_{\rm m}(n)} \frac{n^{(k+1)t_{\rm m}(n)}}{((k+1)t_{\rm m}(n))!} (2\sqrt{k}+W)^n \left(\frac{2\sqrt{k}}{2\sqrt{k}+W} \right)^{2r_{\rm m}(n)+2} \ . \tag{93}$$

Using standard bounds on factorials and binomial coefficients and the fact that $t_{\rm m} \ll r_{\rm m}$, we can conclude on the existence of $\mu > 0$ such that for n large enough

$$\frac{1}{d_n^0} \sum_{\omega \in \mathcal{W}_n^2} \pi^0(\omega) \leq e^{-\mu \frac{n}{(\ln n)^{2+x}}} \ . \tag{94}$$

Finally, putting together (79), (87) and (94), we obtain for n large enough:

$$\frac{d_n}{d_n^0} \leq \left(\mathbb{E}\left[\frac{V}{W} \right] \right)^{\lfloor \frac{n}{(\ln n)^{2+2x}} \rfloor} + e^{-\gamma \frac{n}{(\ln n)^2}} + e^{-\mu \frac{n}{(\ln n)^{2+x}}} \ . \tag{95}$$

The first term dominates when $n \to \infty$, hence we conclude that (renaming $2x$ in x) for all $x > 0$ and n large enough :

$$\frac{1}{n} \ln(d_n) \leq \ln(2\sqrt{k}+W) - \frac{1}{(\ln n)^{2+x}} \ln(\mathbb{E}[V/W]^{-1}) \ . \tag{96}$$

The same remarks as those following (44) hold, namely that the positive constant $\ln(\mathbb{E}[V/W]^{-1})$ is irrelevant because of the condition $x > 0$, and that an improvement of the proof would be in any case limited by the requirement $\gamma < \frac{(\pi \ln k)^2}{2} \frac{2\sqrt{k}}{2\sqrt{k}+W}$, in agreement with the lower bound in Eq. (12).

VI. CONCLUSIONS

In this paper we have characterized the double-exponential Lifshitz behavior of the Bethe lattice density of states close to its edge in presence of bounded (diagonal or off-diagonal) disorder in an indirect way, by controlling the asymptotic growth of its moments. As a byproduct of this study we have unveiled some geometric properties of the dominant supports of closed random walks on regular trees. Let us conclude by making a series of observations and suggestions for future work.

We believe that most of the methods and results of the paper could be adapted to characterize the integrated density of states of the adjacency matrix of random graphs with arbitrary bounded degree distribution. Denoting $k_{\rm max}$ the maximum degree, one expects a Lifshitz tail phenomenon to occur around $2\sqrt{k_{\rm max}}$ as soon as the random graph is not $k_{\rm max}$-regular. To characterize it one could use the results of the off-diagonal disorder model on $\mathbb{T}_{k_{\rm max}-1}$, with $J_{ij} \in \{0,1\}$. The J_{ij} are i.i.d. only in the case of percolation (i.e. for a binomial degree distribution), but in the general case two J_{ij} are correlated only when they share a common vertex, so that it should be possible to handle these weak correlations, thus extending the results of [35, 36].

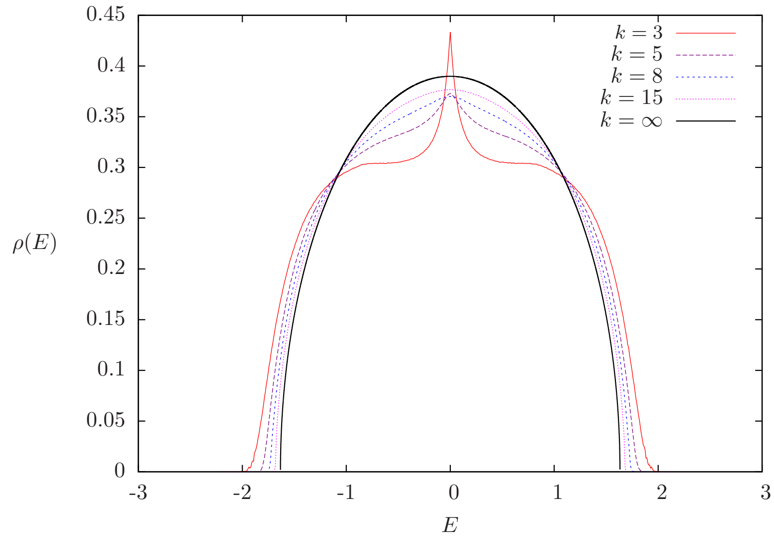


FIG. 9: The density of states in presence of off-diagonal disorder for different degrees, with rescaled couplings J_{ij} uniformly random on $[-\sqrt{2/k}, \sqrt{2/k}]$. The $k = \infty$ curve corresponds to the Wigner semi-circle law of support $[-\sqrt{8/3}, \sqrt{8/3}]$.

When looking at the plot of the off-diagonal disorder density of states in Fig. 1 one could be tempted to divide its support in two regions, one where it is "large" and one where the Lifshitz phenomenon strongly reduces it. However the density of states is strictly positive everywhere inside its support, so that there is no unambiguous criterion on the location of the frontier between these two regimes. Such a distinction can be sharply defined by letting another parameter of the model evolve, i.e. the degree $k+1$ of the lattice. In Fig. 9 we have plotted the density of states of the off-diagonal disordered model for various values of k; to allow for a meaningful comparison of these curves we have scaled in accordance the magnitude of the disorder, taken to be uniform on $[-\sqrt{2/k}, \sqrt{2/k}]$. The support of the density of states is thus, independently of k, the interval $[-2\sqrt{2}, 2\sqrt{2}]$. However the density of states converges (pointwise) in this limit to the Wigner semi-circle law whose support is controlled by the variance of the J_{ij} (and not by their maximal value as the support of the density of states), and which is found here to be $[-\sqrt{8/3}, \sqrt{8/3}]$. The regime $|E| \in [\sqrt{8/3}, 2\sqrt{2}]$ is thus, in this $k \to \infty$ limit, the one of the Lifshitz tail. Note the similarity with the unbounded (Gaussian or Cauchy) diagonal disorder case studied in [16, 21, 26, 27], with $1/k$ playing here the same role as the multiplicative constant in front of the unbounded V_i's in these studies.

A possible direction for future studies would be the investigation of the distribution of the largest eigenvalue of the Anderson model defined on random regular graphs, more precisely of its deviation from the upper limit of the density of states. In the Gaussian ensembles of random matrix theory the density of states is the Wigner semi-circle law, which vanishes as a square root at its edge; in that case the typical fluctuations of the largest eigenvalue around the edge of the density of states are of order $N^{-2/3}$ and described by the Tracy-Widom law [53]. This result has been extended by Sodin to the case of random regular graphs of fixed degree with $J_{ij} = \pm 1$ [54, 55]. In presence of a Lifhsitz tail in the density of states both the scaling with N and the distribution of the fluctuations of the largest eigenvalue should be modified.

Finally, another direction of investigation could concern the numerical procedures used to solve Recursive Distributional Equations as (8). The sample representation at the basis of the population dynamics [12, 40] is very natural and allows simple and versatile implementations of the method, yet it performs poorly in sampling rare events of the Lifshitz tail type. A combination between a sample representation of the typical part of the distribution and a mesh representation for its very small probability part might provide a better alternative in such cases.

Acknowledgments

This work has been partly funded by the ANR project BOOLE, ANR-09-BLAN6011. We warmly thank Giulio Biroli, Charles Bordenave, Marc Lelarge, Justin Salez, Sasha Sodin and Marco Tarzia for very useful discussions.

Appendix A: Numerical determination of the density of states and the center of the band singularity

Let us first explain the numerical procedure we followed to obtain the density of states displayed in Figs. 1,3 and 9. As shown in Sec. II B this computation amounts to solve the Recursive Distributional Equation (RDE) of Eq. (8), and then to compute the average given in (9). The numerical procedure we implemented to solve this problem, known as population dynamics [12, 40] or pool method [18], consists in approximating the distribution of $\widetilde{G}$ by the empirical distribution over a sample of $\mathcal{N} \gg 1$ (in practice we used $\mathcal{N}$ of the order of 10^4) representative values of $\widetilde{G}$. This sample is updated according to Eq. (8) until numerical convergence is reached, then the average in (9) is estimated as an empirical average over the sample (and over several update steps, a few hundreds for the data shown, to increase the precision). The numerical accuracy of this method is controlled by the size $\mathcal{N}$ of the sample. The limit on the accessible values of $\mathcal{N}$ put by the memory available on present computers is such that the regime of Lifshitz tails, where the density of states is extremely small and hence requires a huge precision to be determined, is not reachable via such a numerical procedure.

We shall now present the heuristic arguments which led us to the conjecture that in the case of continuous off-diagonal disorder with support intersecting a neighbourhood of 0, the density of states exhibit a singularity at $E = 0$ of the form

$$\rho(E) \sim \rho(0) - \alpha |E|^{\frac{k-1}{2}} \quad \text{as} \quad E \to 0 \ , \tag{A1}$$

where α is a positive constant. This conjecture is in very good agreement with our numerical results for $k = 2, 3, 4$, as shown explicitly for $k = 2$ in the right panel of Fig. 3.

Consider the RDE (8) on $\widetilde{G}$ at the center of the band ($E = 0$), and in the limit of vanishing regularizing imaginary part η (a similar limit of the RDE has been studied rigorously in [56], with $J_i = 1$ and k a Poisson random variable). To simplify notations we introduce the random variable $\widetilde{g} \equiv -i\widetilde{G}(0)$, solution of the RDE

$$\widetilde{g} \overset{\mathrm{d}}{=} \frac{1}{\sum_{i=1}^{k} J_i^2 \widetilde{g}_i} \ . \tag{A2}$$

Since for $\operatorname{Im} z > 0$ the relevant solution of the RDE (8) has $\operatorname{Im} \widetilde{G}(z) > 0$, we shall look for a solution of (A2) supported on $[0, \infty[$. We will assume that $\widetilde{g}$ has a density $\widetilde{\mu}$ that decays as a power law for large values of $\widetilde{g}$,

$$\widetilde{\mu}(\widetilde{g}) \underset{\widetilde{g} \to \infty}{\simeq} \widetilde{g}^{-x-1} \ , \tag{A3}$$

where in this appendix $\simeq$ means asymptotic equivalence within a multiplicative constant, and x is an exponent that we shall determine self-consistently. First note that (A3) is equivalent to

$$\mathbb{P}(\widetilde{g} \geq A) \underset{A \to \infty}{\simeq} \frac{1}{A^x} \ . \tag{A4}$$

The recursive distributional equation (8) relates the assumption on the behavior of $\widetilde{\mu}$ near ∞ to its behavior near 0:

$$\begin{aligned} \mathbb{P}\left(\widetilde{g} \leq a\right) &= \mathbb{P}\left(\sum_{i=1}^{k} J_i^2 \widetilde{g}_i \geq \frac{1}{a} \right) \\ &\underset{a \to 0}{\simeq} \mathbb{P}\left(J^2 \widetilde{g} \geq \frac{1}{a} \right) \\ &\underset{a \to 0}{\simeq} \mathbb{P}\left(\widetilde{g} \geq \frac{1}{a} \right) \underset{a \to 0}{\simeq} a^x \ . \end{aligned} \tag{A5}$$

In the first step we have used the heavy tail character of $\widetilde{g}$, hence the probability that the sum is large is close to the probability that a single term is large, the second step relies on the boundedness of J, and the last one follows from

(A4). Now we use again (8) in the other direction, i.e. to deduce the behavior at large $\widetilde{g}$ from the one at small $\widetilde{g}$:

$$\begin{aligned}\mathbb{P}(\widetilde{g} \geq A) &= \mathbb{P}\left(\sum_{i=1}^{k} J_i^2 \widetilde{g}_i \leq \frac{1}{A}\right) \\ &\underset{A\to\infty}{\simeq} \left[\mathbb{P}\left(J^2 \widetilde{g} \leq \frac{1}{A}\right)\right]^k \\ &\underset{A\to\infty}{\simeq} \left[\mathbb{P}\left(J^2 \leq \frac{1}{A}\right)\right]^k \\ &\underset{A\to\infty}{\simeq} \left[\frac{1}{\sqrt{A}}\right]^k .\end{aligned} \tag{A6}$$

Indeed the probability that the sum in the first line is small is roughly the probability that all the (positive) terms are small. In the second step we assumed $x > 1/2$, hence the probability that the product $J^2\widetilde{g}$ is small is controlled by the smallness of J^2, and the last step follows from the assumption that J has a finite density near 0 (note that if the support of J were bounded away from 0 there would not be any singularity in the density of states). Comparing this with (A4) gives the self-consistency condition $x = k/2$, which confirms the hypothesis $x > 1/2$ for all relevant k. The same computations applied to the density μ of $g \equiv -iG(0)$ yield the expansion:

$$\mu(g) \underset{g\to\infty}{\simeq} g^{-\frac{k+1}{2}-1} \sim \alpha' \, g^{-\frac{k+1}{2}-1} \qquad (\alpha' \in \mathbb{R}_+) . \tag{A7}$$

Now we reintroduce the regularizing imaginary part η, and observe that when it is small its main effect is to provide a cutoff at $1/\eta$ on the distribution of $\widetilde{g}$ and g determined above directly at $\eta = 0$. This yields

$$\begin{aligned}\frac{1}{\pi}\mathrm{Im}\mathbb{E}[G(0+i\eta)] &\underset{\eta\to 0}{\simeq} \frac{1}{\pi}\int_0^{1/\eta} g\mu(g)\mathrm{d}g \\ &= \frac{1}{\pi}\int_0^{\infty} g\mu(g)\mathrm{d}g - \frac{\alpha'}{\pi}\int_{1/\eta}^{\infty} g^{-\frac{k+1}{2}}\left(1+o(1)\right)\mathrm{d}g = \rho(0) - \alpha\, \eta^{\frac{k-1}{2}} + o\left(\eta^{\frac{k-1}{2}}\right) .\end{aligned} \tag{A8}$$

The finiteness of $\rho(0)$ follows from the integrability of $g\mu(g)$, and α is a positive constant proportional to α'. If we assume finally that we can make the substitution $\eta \to -iE$ in this expression, and that it corresponds to $\eta^{\frac{k-1}{2}} \to |\eta|^{\frac{k-1}{2}} \to |E|^{\frac{k-1}{2}}$, we get:

$$\rho(E) = \rho(0) - \alpha|E|^{\frac{k-1}{2}} + o(E^{\frac{k-1}{2}}) , \tag{A9}$$

a form in perfect agreement with the numerical determination of the density of states for $k = 2, 3$ and 4 (shown in Fig. 3 for $k = 2$). The justification of the jump from (A8) to (A9) is not obvious; indeed, the small η expansion shows that ρ is not analytic near 0, hence no 'analytic continuation' argument can be straightforwardly applied. However as the density of states has to be invariant under $E + i\eta \to -E + i\eta$, one could assume for $\mathrm{Im}z > 0$ an Ansatz $\rho(z) = \rho(0) + f_1(|z|) + f_2(z)$ with $f_{1,2}$ regular and $f_2 = o(f_1)$ when $|z| \to 0$.

Appendix B: From the moments of a random variable to its cumulative distribution function

In Sec. II D we have seen how the asymptotic behavior of a probability density close to the edge of its support translates into the behavior of its large order moments. This appendix is devoted to the reverse direction of this connection, namely to what can be learnt on a probability measure from the asymptotic knowledge of its moments. In particular we provide the proofs of Eqs. (14,15).

Consider a non-negative random variable X, with a probability distribution η supported on $[0, E_+]$, where E_+ is a priori unknown. Note that the assumption that η is supported on non-negative reals is not restrictive for the cases we want to consider here: in the off-diagonal disorder case one can use the symmetry of ρ to come back to the above situation, and in the diagonal case the density of states is supported on $[E_-, E_+]$ with $|E_-| < |E_+|$, hence the negatively supported part of the probability measure is asymptotically irrelevant. Thus Eqs. (14,15) will be direct consequences of (11,12) and (B7).

Let us decompose the expression of the n-th moment of X, for an arbitrary $\delta \in [0, E_+]$, as:

$$u_n = \mathbb{E}[X^n] = \mathbb{E}[X^n | X < E_+ - \delta]\mathbb{P}[X < E_+ - \delta] + \mathbb{E}[X^n | X \geq E_+ - \delta]\mathbb{P}[X \geq E_+ - \delta] . \tag{B1}$$

From that equation one easily obtains:

$$(1-N(E_+-\delta))\,(E_+-\delta)^n \leq u_n \leq (E_+-\delta)^n + E_+^n(1-N(E_+-\delta)) \ . \tag{B2}$$

Hence one has, for arbitrary n:

$$1-N(E_+-\delta) \leq (E_+-\delta)^{-n} u_n \ , \tag{B3}$$

$$1-N(E_+-\delta) \geq E_+^{-n} u_n - \left(1-\frac{\delta}{E_+}\right)^n \ . \tag{B4}$$

As this is valid for any $\delta > 0$, this proves our first point about the link between the support of the density of states and the exponential growth of its moments:

$$\ln E_+ = \lim_{n\to\infty} \frac{1}{n} \ln u_n \ ; \tag{B5}$$

a similar argument can be found in [39].

In order to obtain some bounds on the behaviour of $N(E_+-\delta)$ one has to choose the value of n in Eqs. (B3,B4) in an optimal way, for a given value of δ. As our control on u_n is limited to large values of n, the bounds on the cumulative distribution function shall be relevant only for small values of δ. The result will of course depend on the precise form of u_n, or more precisely on its large n asymptotic. In the following we will assume that u_n has the large n behavior of the form (11,12) of central interest here; the reasoning is however more general and can be applied to reconstruct the behavior of N from u_n in the three examples of Sec. II D. We shall thus make the following assumption on the large n behaviour of u_n in order to treat within the same frame the off-diagonal and diagonal disorder cases:

Assume that for some $c, \epsilon, x > 0$ and $n_0 \in \mathbb{N}$ it holds for $n \geq n_0$ that

$$-(1+\epsilon)E_+\frac{c}{(\ln n)^2} \leq \frac{1}{n}\ln u_n - \ln(E_+) \leq -E_+\frac{c}{(\ln n)^{2+x}} \ . \tag{B6}$$

Then for $\delta > 0$ small enough one has:

$$e^{-e^{\sqrt{\frac{E_+^2 c(1+3\epsilon)}{\delta}}}} \leq 1-N(E_+-\delta) \leq e^{-e^{\left(\frac{E_+^2 c}{\delta}\right)^{\frac{1}{2+2x}}}} \ . \tag{B7}$$

We start with the right inequality. Using the upperbound on u_n from (B6), (B3) reads for $n \geq n_0$:

$$1-N(E_+-\delta) \leq \left(1-\frac{\delta}{E_+}\right)^{-n} e^{-E_+ c\frac{n}{(\ln n)^{2+x}}} \ . \tag{B8}$$

Now we take:

$$n = \lfloor e^{\left(\frac{E_+^2 c}{\delta}\right)^{\frac{1}{2+x}}} \rfloor + 1 \ . \tag{B9}$$

For δ small enough n becomes larger than n_0 and one gets, replacing into (B8):

$$1-N(E_+-\delta) \leq e^{-e^{\left(\frac{E_+^2 c}{\delta}\right)^{\frac{1}{2+2x}}}} \ . \tag{B10}$$

The left inequality can be obtained in a similar way: this time (B4) may be rewritten, under assumption (B6) and for n large enough:

$$1-N(E_+-\delta) \geq e^{-(1+\epsilon)E_+ c\frac{n}{(\ln n)^2}} - \left(1-\frac{\delta}{E_+}\right)^n \ . \tag{B11}$$

To obtain a valid bound we want the first term to be much larger than the second, that is:

$$e^{-(1+\epsilon)E_+ c\frac{1}{(\ln n)^2}} > \left(1-\frac{\delta}{E_+}\right) \ . \tag{B12}$$

Thus we take

$$n = \lfloor e^{\sqrt{\frac{E_+^2 c(1+2\epsilon)}{\delta}}} \rfloor \tag{B13}$$

which satisfies (B12) and gives, replacing into (B11):

$$1-N(E_+-\delta) \geq e^{-e^{\sqrt{\frac{E_+^2 c(1+3\epsilon)}{\delta}}}} \ . \tag{B14}$$

Appendix C: Proof of the recursion formula for the moments

In this appendix we give a formal proof of Eqs. (20,21), making use of generating functions; we follow the standard notations in this field, namely if $F(z,x)$ is a formal series, $[z^n x^q]F(z,x)$ denotes the coefficient of its $z^n x^q$ term.

Let us first introduce, for a given realization of the disorder, the generating function $F(z) = \langle 0|\frac{1}{\mathbb{I}-zH}|0\rangle$, which is equal up to a change of variables to the resolvent at the root of $\mathbb{T}_k$. We thus have $u_n = \mathbb{E}[[z^n]F(z)]$. We decompose the operator H according to

$$H = V|0\rangle\langle 0| + \sum_{i=1}^{k+1} J_i \left(|i\rangle\langle 0| + |0\rangle\langle i|\right) + \sum_{i=1}^{k+1} \widetilde{H}_i \ , \tag{C1}$$

where $\widetilde{H}_i$ acts only on the subtree rooted at i, one of the $k+1$ neighbors of the root. We introduce the generating functions for each of these subtrees, $F_i(z) = \langle 0|\frac{1}{\mathbb{I}-z\widetilde{H}_i}|0\rangle$, and use the resolvent identity between operators $\frac{1}{A} - \frac{1}{B} = \frac{1}{A}(B-A)\frac{1}{B}$ with $A = \mathbb{I} - zH$, $B = \mathbb{I} - z\left(V|0\rangle\langle 0| + \sum_{i=1}^{k+1}\widetilde{H}_i\right)$, to get:

$$\begin{aligned} F(z) &= \frac{1}{1-zV} + \left\langle 0 \left| \frac{z}{\mathbb{I}-zH} \sum_{i=1}^{k+1} J_i \left(|i\rangle\langle 0| + |0\rangle\langle i|\right) \frac{1}{\mathbb{I} - zV|0\rangle\langle 0| - z\sum_{i=1}^{k+1}\widetilde{H}_i} \right| 0 \right\rangle \\ &= \frac{1}{1-zV}\left(1 + z\sum_{i=1}^{k+1} J_i \left\langle 0 \left| \frac{1}{\mathbb{I}-zH} \right| i \right\rangle\right) \ , \end{aligned} \tag{C2}$$

where we used the fact that $\widetilde{H}_i|0\rangle = 0$. One can show in a similar way that $\left\langle 0 \left| \frac{1}{\mathbb{I}-zH} \right| i \right\rangle = zJ_iF(z)F_i(z)$, so that

$$F(z) = \frac{1}{1-zV}\left(1 + F(z)z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)\right) \Longrightarrow F(z) = \frac{1}{1 - zV - z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)} \ . \tag{C3}$$

Let us now introduce a bivariate generating function, $F(z,x) = \frac{1}{1-xF(z)}$. As a consequence of the recursion equation on $F(z)$, one obtains:

$$\begin{aligned} F(z,x) &= \frac{1 - zV - z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)}{1 - x - zV - z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)} \\ &= \frac{1}{1-x}\left(1 - zV - z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)\right)\sum_{l=0}^{\infty}\left(\frac{zV + z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)}{1-x}\right)^l \\ &= \frac{1}{1-x} + \sum_{l=1}^{\infty}\frac{1}{(1-x)^l}\left(\frac{1}{1-x} - 1\right)\left(zV + z^2\sum_{i=1}^{k+1} J_i^2 F_i(z)\right)^l \\ &= \frac{1}{1-x} + \sum_{\substack{p,s\geq 0 \\ p+s\geq 1}} \frac{x}{(1-x)^{s+p+1}} z^{2p+s} \sum_{\substack{p_1,\dots,p_{k+1}\geq 0 \\ p_1+\dots+p_{k+1}=p}} \binom{s+p}{s, p_1, \dots, p_{k+1}} V^s \prod_{i=1}^{k+1} J_i^{2p_i} \prod_{i=1}^{k+1} F_i(z)^{p_i} \ . \end{aligned} \tag{C4}$$

As $u_n = \mathbb{E}[[z^n x]F(z,x)]$, we obtain for $n \geq 1$

$$u_n = \sum_{\substack{s,p_1,m_1,\dots,p_{k+1},m_{k+1}\geq 0 \\ s+m_1+2p_1+\dots+m_{k+1}+2p_{k+1}=n}} \binom{s+p}{s, p_1, \dots, p_{k+1}} \mathbb{E}[J^{2p_1}]\dots\mathbb{E}[J^{2p_{k+1}}]\mathbb{E}[V^s] \prod_{i=1}^{k+1} \mathbb{E}[[z^{m_i}]F_i(z)^{p_i}] \ , \tag{C5}$$

which proves Eq. (20) with $u(n,q) = \mathbb{E}[[z^n]F_i(z)^q] = \mathbb{E}[[z^n x^q]F_i(z,x)]$. In this last expression we introduced the bivariate generating function $F_i(z,x) = \frac{1}{1-xF_i(z)}$. The proof of Eq. (21) follows exactly the same lines, using $\widetilde{\mathbb{T}}_k$ instead of $\mathbb{T}_k$ (hence the root has only k neighbors), and extracting the coefficient of order x^q in the equation corresponding to (C4) thanks to the identity $[x^{q-1}](1-x)^{-(s+p+1)} = \binom{s+p+q-1}{q-1}$.

Appendix D: Computation of $u(n,q)$ in the pure case

In this appendix we explain how to find the solution (25) to the recursive equation (21) in the pure case. Following the interpretation of $u(n,q)$ in terms of walks on $\widetilde{\mathbb{T}}_k \cup (-1)$ given in Sec. III B 1, one realizes that in the pure case $u^0(n,q)$ is $k^{n/2}$ times the number of Dyck paths of length $n+2q$ that start and end at height 1 and visit q times the height 0. Following the reflection principle discussed in [50], this amounts to count the number of paths from $(0,0)$ to (n,q) that reach the height q for the first time at abscissa n, hence $u^0(n,q) = \binom{n+q}{n/2+q}\frac{q}{n+q}k^{n/2}$. In particular this gives back the well-known value of $\tilde{c}^0_n = u^0(n,1)$ as the Catalan number times $k^{n/2}$.

One can also check analytically that this expression for $u^0(n,q)$ is indeed the solution of the recursion relation (21). The simplest way to do so is to take benefit of the equivalence between (21) and the equations on bivariate generating functions from which we derived (21) in App. C. Simplifying them in the pure case, one realizes that the verification boils down to prove:

$$\binom{n+q}{n/2+q}\frac{q}{n+q}k^{n/2} = [z^n]\left(\frac{1-\sqrt{1-4kz^2}}{2kz^2}\right)^q \ . \tag{D1}$$

This equality can be proven [50] by checking it directly for $q=1$ and all n, and by showing that both sides $l_{n,q}$ and $r_{n,q}$ of the equation obeys the same recurrence equation,

$$l_{n,q} = \frac{l_{n+2,q-1} - l_{n+2,q-2}}{k} \ , \qquad r_{n,q} = \frac{r_{n+2,q-1} - r_{n+2,q-2}}{k} \ . \tag{D2}$$

Appendix E: Dyck and Motzkin paths of restricted height

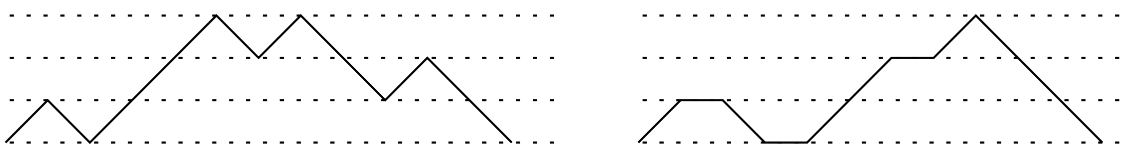


FIG. 10: Left: a Dyck path of length 12 and height 3. Right: a Motzkin path of length 11 and height 3.

We explain in this appendix the derivation of the properties of Motzkin paths of restricted height that we used during the proofs of the lower bounds in Sec. IV A and V A. A Motzkin path of length n is a sequence $h_0, h_1, \dots, h_n$ of non-negative integers such that $h_0 = h_n = 0$, $h_{i+1} - h_i \in \{+1, 0, -1\}$ for all $i \in [0, n-1]$, see Fig. 10 for an illustration. It is called a Dyck path if there is no horizontal step $h_{i+1} = h_i$. Its height is the maximum value of h_i reached for $i \in [0,n]$. Let us assign a weight a (resp. c, 1) to an ascending (resp. horizontal, descending) step, and define the weight of a Motzkin path as the product of the weight of its steps. We denote by $m_{n,h}$ the sum of these weights over the paths of length n whose height is smaller or equal to h. The generating function $G^{(h)}(x) = \sum_{n\geq 0} m_{n,h} x^n$ is known to be [51, 52]:

$$G^{(h)}(x) = \frac{1}{x}\frac{p_h(1/x)}{p_{h+1}(1/x)} \ , \tag{E1}$$

where the $p_h(x)$ are polynomials of degree h, solution of the recurrence equation:

$$p_{h+1}(x) = (x-c)\,p_h(x) - a\,p_{h-1}(x) \ , \qquad \text{with} \quad p_0(x) = 1 \ , \quad p_1(x) = x - c \ . \tag{E2}$$

It is convenient to change variables and define the polynomials q_h by $q_h(x) = \frac{1}{a^{r/2}}p_h(2\sqrt{a}\,x + c)$. Indeed the recursion becomes:

$$q_{h+1}(x) = 2x\,q_h(x) - q_{h-1}(x) \ , \qquad \text{with} \quad q_0(x) = 1 \ , \quad q_1(x) = 2x \ . \tag{E3}$$

Hence one recognizes that $q_h(x) = \mathcal{U}_h(x)$, where $\mathcal{U}_h$ is the h-th Chebychev polynomial of the second kind.

Consider now the rational function $\mathcal{U}_h(x)/\mathcal{U}_{h+1}(x)$; it has $h+1$ simple poles located at the roots of $\mathcal{U}_{h+1}(x)$, namely $x_j(h+1) = \cos\left(\frac{j\pi}{h+2}\right)$ with $j \in [1, h+1]$. Using basic properties of the Chebychev polynomials one easily obtains the

following decomposition:

$$\frac{\mathcal{U}_h(x)}{\mathcal{U}_{h+1}(x)} = \sum_{j=1}^{h+1} \frac{1}{h+2} \sin^2\left(\frac{j\pi}{h+2}\right) \frac{1}{x - x_j(h+1)} \ . \tag{E4}$$

Replacing the p_h in terms of the Chebychev polynomials in Eq. (E1) and using the decomposition above leads to an explicit expression of the generating function,

$$G^{(h)}(x) = \sum_{j=1}^{h+1} \frac{2}{h+2} \sin^2\left(\frac{j\pi}{h+2}\right) \frac{1}{1 - x\left(c + 2\sqrt{a}\cos\left(\frac{j\pi}{h+2}\right)\right)} \ , \tag{E5}$$

and of the coefficient $m_{n,h}$ after the expansion in powers of x,

$$m_{n,h} = \sum_{j=1}^{h+1} \frac{2}{h+2} \sin^2\left(\frac{j\pi}{h+2}\right) \left(c + 2\sqrt{a}\cos\left(\frac{j\pi}{h+2}\right)\right)^n \ . \tag{E6}$$

This expression is exact for all n and h. For n even all terms in the sum are positive and one can lowerbound it by retaining only the term $j = 1$ and by using the inequality $\sin\theta \geq 2\theta/\pi$ for $\theta \in [0, \pi/2]$. This yields the bound

$$m_{n,h} \geq \left(\frac{2}{h+2}\right)^3 \left(c + 2\sqrt{a}\cos\left(\frac{\pi}{h+2}\right)\right)^n \ , \tag{E7}$$

which we used in Sec. IV A with $a = 1$ and $c = 0$.

Let us now justify the asymptotic statement on $m_{n,h}$ used in Sec. V A . Consider first the behavior of $m_{n,h}$ for $c > 0$, $n \to \infty$ with h fixed. In Eq. (E6) the term raised to the power n is maximal for $j = 1$, and is strictly greater in absolute value than all others, hence

$$\lim_{n\to\infty} \frac{1}{n} \ln m_{n,h} = \ln\left(c + 2\sqrt{a}\cos\left(\frac{\pi}{h+2}\right)\right) \ . \tag{E8}$$

Suppose now that h diverges in an n-dependent way. As long as $h \ll n^{1/2}$ one can check that the terms $j \geq 2$ are subdominant, and that

$$\frac{1}{n} \ln m_{n,h} = \ln(c + 2\sqrt{a}) - \frac{\pi^2\sqrt{a}}{c + 2\sqrt{a}} \frac{1}{h^2} + O\left(\frac{1}{h^4}\right) \ . \tag{E9}$$

This is the result used in Sec. V A with $a = k$, $c = V_0$ and $h = \lfloor \alpha \ln n \rfloor$, which indeed satisfied the condition $h \ll n^{1/2}$.

Appendix F: The degeneracy of walks induced by self-bond steps

This appendix is devoted to an alternative proof a statement used in Sec. V B, namely that there are $\binom{n}{s}$ walks ω of length n with s self-bond steps that share a common underlying walk $\overline{\omega}$ of length $n - s$, obtained by removing the self-bond steps from ω. Let us denote $\widehat{\sigma}$ the common skeleton of length $n - s$. From Eq. (22) one reads the combinatorial factor associated to such a skeleton complemented by the numbers $\{s_v\}_{v\in\mathcal{V}}$ of self-bond steps on each of the vertices:

$$\eta(\widehat{\sigma}, \{s_v\}) = \binom{s_0 + n_1 + n_2 + \cdots + n_{k+1}}{s_0, n_1, n_2, \ldots, n_{k+1}} \prod_{v\in\mathcal{V}\setminus 0} \binom{n_v - 1 + s_v + n_{v_1} + \cdots + n_{v_k}}{n_v - 1, s_v, n_{v_1}, \ldots, n_{v_k}} \tag{F1}$$

$$= \eta(\widehat{\sigma}) \binom{s_0 + n_1 + n_2 + \cdots + n_{k+1}}{s_0} \prod_{v\in\mathcal{V}\setminus 0} \binom{n_v - 1 + s_v + n_{v_1} + \cdots + n_{v_k}}{s_v} \tag{F2}$$

In the second line we have factored out the combinatorial factor $\eta(\widehat{\sigma})$ of the walk deprived of the self-bond steps. We compute now the sum over all possible choices of the $\{s_v\}_{v\in\mathcal{V}}$:

$$\begin{aligned}\sum_{\substack{\{s_v\}_{v\in\mathcal{V}}\\ \sum_v s_v=s}} \eta(\widehat{\sigma},\{s_v\}) &= \eta(\widehat{\sigma})[z^s]\left(\sum_{s_0=0}^{\infty}\binom{s_0+n_1+n_2+\cdots+n_{k+1}}{s_0}z^{s_0}\right)\prod_{v\in\mathcal{V}\setminus 0}\left(\sum_{s_v=0}^{\infty}\binom{n_v-1+s_v+n_{v_1}+\cdots+n_{v_k}}{s_v}z^{s_v}\right)\\ &= \eta(\widehat{\sigma})[z^s](1-z)^{-(1+n_1+n_2+\cdots+n_{k+1})}\prod_{v\in\mathcal{V}\setminus 0}(1-z)^{-(n_v+n_{v_1}+\cdots+n_{v_k})} \qquad \text{(F3)}\\ &= \eta(\widehat{\sigma})[z^s](1-z)^{-(1+2\sum_{v\in\mathcal{V}\setminus 0}n_v)}=\eta(\widehat{\sigma})\binom{n}{s}\ ,\end{aligned}$$

where we have used the combinatorial identity $[z^p](1-z)^{-(a+1)}=\binom{p+a}{p}$, recognized that for each edge e n_e appears twice (in the binomial coefficient associated to both its endvertices), and noted that by definition of the length of the skeleton $2\sum_{v\in\mathcal{V}\setminus 0}n_v=n-s$. Each of the $\eta(\widehat{\sigma})$ walks ϖ without self-bond steps compatible with $\widehat{\sigma}$ gives thus birth to $\binom{n}{s}$ walks with s self-bond steps.

We shall now explain the proof of the upperbound (89) on $E(\varpi,s,\mathcal{A})$, the number of walks $\omega\in\mathcal{W}_n$ that are obtained from $\varpi\in\mathcal{M}_{n-s}$ by adding to it s self-bond steps, in such a way that the self-support of ω is $\mathcal{A}$. This number is smaller than the number of walks ω with a self-support included in $\mathcal{A}$ (not necessarily equal to $\mathcal{A}$). This last quantity can be computed as

$$\frac{1}{\eta(\widehat{\sigma})}\sum_{\substack{\{s_v\}_{v\in\mathcal{A}}\\ \sum_v s_v=s}}\eta(\widehat{\sigma},\{s_v\})=\binom{s-1+\mathbb{I}(0\in\mathcal{A})(1+n_1+\cdots+n_{k+1})+\sum_{v\in\mathcal{A}\setminus 0}(n_v+n_{v_1}+\cdots+n_{v_k})}{s}\ , \qquad \text{(F4)}$$

where we denoted $\mathbb{I}(M)$ the characteristic function of the event M, and obtained the right-hand side with a reasoning similar to the one in (F3). One can trade the sum over vertices for a sum over the edges of the support by introducing the numbers $d_e(\mathcal{A})\in\{0,1,2\}$ which count the number of endvertices of the edge e which belongs to $\mathcal{A}$, and rewrite the last quantity as

$$\binom{n-(\mathbb{I}(0\notin\mathcal{A})+\sum_{e\in\sigma}n_e(\varpi)(2-d_e(\mathcal{A})))}{s}\ . \qquad \text{(F5)}$$

Finally (89) is obtained by using the facts that $n_e(\varpi)\geq 1$ for all edges in the support and that each vertex of $\mathcal{A}$ has at most degree $k+1$ in the support.

[1] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[2] E. Abrahams, editor, *50 years of Anderson localization* (World Scientific, 2010).
[3] R. Carmona and J. Lacroix, *Spectral Theory of Random Schrödinger Operators* (Birkhäuser, 1990).
[4] L. A. Pastur and A. Figotin, *Spectra of Random and Almost-Periodic Operators* (Springer-Verlag, 1992).
[5] P. Stollmann, *Caught by disorder* (Birkhäuser, 2001).
[6] I. Veselić, *Existence and Regularity Properties of the Integrated Density of States of Random Schrödinger Operators* (Springer-Verlag, 2007).
[7] F. Wegner, Z. Phys. B **44**, 9 (1981).
[8] B. Simon, Phys. Rev. B **27**, 3859 (1983).
[9] I. M. Lifshitz, Advances in Physics **13**, 483 (1964).
[10] B. Simon, J. Stat. Phys. **38**, 65 (1985).
[11] J. Fröhlich and T. Spencer, Communications in Mathematical Physics **88**, 151 (1983).
[12] R. Abou-Chacra, D. J. Thouless, and P. W. Anderson, Journal of Physics C: Solid State Physics **6**, 1734 (1973).
[13] R. Abou-Chacra and D. J. Thouless, Journal of Physics C: Solid State Physics **7**, 65 (1974).
[14] Y. Kim and A. B. Harris, Phys. Rev. B **31**, 7393 (1985).
[15] A. D. Mirlin and Y. V. Fyodorov, Nuclear Physics B **366**, 507 (1991).
[16] J. Miller and B. Derrida, Journal of Statistical Physics **75**, 357 (1994).
[17] F.X. Bronold, A. Alvermann, and H. Fehske, Phil. Mag. **84**, 673 (2004).
[18] C. Monthus and T. Garel, Journal of Physics A: Mathematical and Theoretical **42**, 075002 (2009).
[19] G. Biroli, G. Semerjian, and M. Tarzia, Progress of Theoretical Physics Supplement **184**, 187 (2010).

[20] H. Kunz and B. Souillard, J. Physique Lett. **44**, 411 (1983).
[21] V. Acosta and A. Klein, Journal of Statistical Physics **69**, 277 (1992).
[22] M. Aizenman and S. Molchanov, Communications in Mathematical Physics **157**, 245 (1993).
[23] A. Klein, Advances in Mathematics **133**, 163 (1998).
[24] M. Aizenman, R. Sims, and S. Warzel, Probability Theory and Related Fields **136**, 363 (2006).
[25] R. Froese, D. Hasler, and W. Spitzer, Communications in Mathematical Physics **269**, 239 (2007).
[26] M. Aizenman and S. Warzel, Physical Review Letters **106**, 136804 (2011).
[27] M. Aizenman and S. Warzel, (2011), arXiv:1104.0969.
[28] G. J. Rodgers and A. J. Bray, Phys. Rev. B **37**, 3557 (1988).
[29] G. Biroli and R. Monasson, Journal of Physics A: Mathematical and General **32**, L255 (1999).
[30] M. Bauer and O. Golinelli, Journal of Statistical Physics **103**, 301 (2001).
[31] G. Semerjian and L. F. Cugliandolo, Journal of Physics A: Mathematical and General **35**, 4837 (2002).
[32] O. Khorunzhy, M. Shcherbina, and V. Vengerovsky, Journal of Mathematical Physics **45**, 1648 (2004).
[33] T. Rogers, I. P. Castillo, R. Kühn, and K. Takeda, Phys. Rev. E **78**, 031116 (2008).
[34] F. L. Metz, I. Neri, and D. Bollé, Phys. Rev. E **82**, 031135 (2010).
[35] P. Müller and P. Stollmann, (2010), arXiv:1002.5000.
[36] T. Reinhold, *Über die integrierte Zustandsdichte des Laplace-Operators auf Bond- Perkolationsgraphen des Bethe-Gitters* (Diploma thesis, 2009), [in German].
[37] O. Khorunzhy, W. Kirsch, P. Müller, and V. Vengerovsky, The Annals of Applied Probability **16**, 295 (2006).
[38] H. Kesten, Trans. Amer. Math. Soc. **92**, 336 (1959).
[39] B. D. McKay, Linear Algebra and its Applications **40**, 203 (1981).
[40] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[41] C. Bordenave and M. Lelarge, Random Structures and Algorithms **37**, 332 (2010).
[42] W. Kirsch, *An Invitation to Random Schrödinger operators* (Panoramas et Syntheses **25**, 1-119, 2008).
[43] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I, Functional Analysis* (Academic Press, 1981).
[44] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, second ed. (John Wiley & Sons, 1971).
[45] H. Kunz and B. Souillard, Communications in Mathematical Physics **78**, 201 (1980).
[46] R. Carmona, A. Klein, and F. Martinelli, Communications in Mathematical Physics **108**, 41 (1987).
[47] F. Klopp, Journal of Mathematical Physics **44**, 4975 (2003).
[48] T. P. Eggarter and R. Riedinger, Phys. Rev. B **18**, 569 (1978).
[49] J.-M. Combes, F. Germinet, and A. Klein, (2010), arXiv:1008.4817.
[50] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, second ed. (John Wiley & Sons, 1971).
[51] P. Flajolet and R. Sedgewick, *Analytic Combinatorics* (Cambridge University Press, 2009).
[52] C. Krattenthaler, Advances in Applied Mathematics **27**, 510 (2001).
[53] C. A. Tracy and H. Widom, Communications in Mathematical Physics **177**, 727 (1996).
[54] S. Sodin, Journal of Statistical Physics **136**, 834 (2009).
[55] S. Sodin, (2009), arXiv:0903.4295v5.
[56] C. Bordenave, M. Lelarge, and J. Salez, Annals of Probability **39**, 1097 (2010).